



Accurate Kidney Tumor Semantic Segmentation

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

Doctor of Philosophy in Science

(Computer Science)

by

Abubaker Abdelrahman

Supervisor

Prof. Serestina Viriri

School of Mathematics, Statistics and Computer Science

UNIVERSITY OF KWAZULU-NATAL

DURBAN-4041 SOUTH AFRICA

EXAMINER'S COPY

December 2024

Preface

The research in this thesis was carried out at the University of KwaZulu-Natal under the supervision of Prof. Serestina Viriri. I hereby declare that all materials incorporated in this thesis are my original work, except where the acknowledgement is made by name or in the form of reference. The work contained herein has not been submitted in part or whole for a degree of any other institution.

Signed A.M.A Abdelrahman Date 06/02/2025

Abubaker Abdelrahman

As the candidate supervisor, I agree to the submission of this dissertation.

Signed  Date 20/02/2025

Prof. Serestina Viriri

Declaration 1 - Plagiarism

I, **Abubaker Abdelrahman** , declare that;

1. The research reported in this thesis, except where otherwise indicated, is my original research.
2. This thesis has not been submitted for any degree or examination at any other university.
3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
 - (a) Their words have been re-written but the general information attributed to them has been referenced,
 - (b) Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.
5. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Signed A.M.A AbdelrahmanDate 06/02/2025

Abubaker Abdelrahman

Declaration 2 - Publications

DETAILS OF CONTRIBUTION TO PUBLICATIONS that form part and/or include research presented in this thesis:

- (i) **A. Abdelrahman, S. Viriri**, “Kidney Tumor Semantic Segmentation Using Deep Learning: A Survey of State-of-the-Art”, *Journal of Imaging*, vol. 8, no. 3:55, (2022). DOI: <https://doi.org/10.3390/jimaging8030055>.
- (ii) **A. Abdelrahman, S. Viriri**, “EfficientNet Family U-Net Models for Deep Learning Semantic Segmentation of Kidney Tumors on CT Images”, *Frontiers in Computer Science*, vol. 5, (2023). DOI: <https://doi.org/10.3389/fcomp.2023.1235622>.
- (iii) **A. Abdelrahman, S. Viriri**, “FPN-SE-ResNet Model for Accurate Diagnosis of Kidney Tumors Using CT Images”, *Applied Sciences*, vol. 13, no. 17: 9802, (2023). DOI: <https://doi.org/10.3390/app13179802>.
- (iv) **A. Abdelrahman, S. Viriri**, Accurate Kidney Tumor Semantic Segmentation Using Deep Learning. *Frontiers in Computer Science, Computer Vision*. [Under review]

Signed A.M.A AbdelrahmanDate 06/02/2025

Abubaker Abdelrahman

Acknowledgements

First and foremost, I thank the Almighty Allah. You have blessed me with the strength to believe in myself and persevere through life's journey. Without my faith in You, the Almighty, none of this would have been possible.

I would like to express my sincere gratitude to my supervisor, Prof. Serestina Viriri, for his continuous support and advice throughout my PhD study and research. His motivation, patience, immense knowledge, and enthusiasm have been invaluable. His guidance helped me a lot during the hard times of research and the writing of this thesis. This thesis reflects the unwavering support and boundless love I received from my family during this challenging academic pursuit. I am immensely grateful to my father and dedicate this thesis to the soul of my mother; may God have mercy on her. Thanks to my small family, dear wife, and sons Mohammed, Munther, Mustafa, and Marwan. God bless and protect them. Thanks to my beloved brother, sisters, and everyone who supported me in completing my academic career. Thanks to my wife's family.

I am also thankful for my friends, who provided both distractions when needed and encouragement when it seemed impossible to continue. The discussions and unwavering support from my peers at the university have enriched my research experience, making this journey not only possible but also enjoyable. Their presence and insights have been pivotal in navigating the complexities of my research topic.

I appreciate the University of Holy Quran and Islamic Sciences, Sudan, for giving me the chance to explore this research work and the University of KwaZulu-Natal for the beautiful and tuition-free research platform they create for research students worldwide

Abstract

Timely and accurate detection of kidney tumors is crucial for improving clinical outcomes, as early diagnosis significantly reduces the risk of progression to advanced stages. Kidney tumors, particularly malignant ones, can be life-threatening if not identified early. Developing a robust, efficient, and automated framework for detecting and classifying kidney tumors from CT images is imperative to enable early intervention and improve survival rates. This study addresses critical challenges in kidney tumor detection, such as variability in tumor characteristics, limited availability of medical experts, and the time-intensive nature of the manual diagnosis. It explores advanced deep learning models to automate and enhance the accuracy and speed of tumor segmentation, offering a solution to bridge the gap in current diagnostic approaches.

The research evaluates the performance of multiple architectures, including the EfficientNet family, SE-ResNet models with Feature Pyramid Network (FPN), and DenseNet integrated with LinkNet. These models were assessed based on key metrics like Intersection over Union (IoU), Dice similarity coefficients, and computational efficiency. EfficientNet-B4 demonstrated exceptional tumor segmentation capabilities, achieving a Tumor Dice score of 0.984 and Mean Dice of 0.990, while EfficientNet-B7 excelled in kidney segmentation with IoU and Dice scores of 0.977 and 0.988, respectively. SE-ResNet50 emerged as the top-performing model in overall segmentation, achieving a Tumor Dice score of 0.992 and a Mean IoU of 0.988. DenseNet121 with LinkNet offered a balance of precision and computational efficiency, achieving Tumor IoU and Dice scores of 0.954 and 0.976, respectively.

This study highlights the potential of integrating scalable deep learning architectures into clinical workflows. EfficientNet models stand out for their resource-constrained adaptability, while SE-ResNet50 demonstrates superior precision, particularly in contexts requiring detailed segmentation. DenseNet121 combines stability and efficiency, offering a versatile solution for varying clinical requirements. The findings underscore the importance of leveraging advanced architectures to enhance kidney cancer diagnostics, paving the way for further optimization and broader real-world deployment. This contribution establishes benchmarks in automated tumor detection and segmentation, providing a foundation for future research aimed at improving diagnostic accuracy and patient outcomes in resource-constrained and high-performance clinical environments.

Contents

Declaration - Supervisor	i
Declaration 1 - Plagiarism	ii
Declaration 2 - Publications	iii
Acknowledgements	iv
Abstract	v
Table of contents	vi
List of figures	vii
List of tables	viii
Abbreviations	x
1 Introduction	1
1.1 Introduction	1
1.2 Motivation	3
1.3 Problem Statement	4
1.4 Thesis Objectives	5
1.5 Thesis Contributions	6
1.6 Thesis Outline	6
1.7 Datasets	7
1.8 Responsible AI	7
2 Literature Review and Related Works : Kidney Tumor Semantic Segmentation Using Deep Learning: A Survey of State-of-the-Art	9
2.1 Introduction	9
2.2 Conclusion	41
3 EfficientNet family U-Net models for deep learning semantic segmentation of kidney tumors on CT images	42
3.1 Introduction	42
3.2 Conclusion	57
4 FPN-SE-ResNet Model for Accurate Diagnosis of Kidney Tumors Using CT Images	58
4.1 Introduction	58

4.2	Conclusion	80
5	Accurate Kidney Tumor Semantic Segmentation Using Deep Learning Techniques on CT Images	81
5.1	Introduction	81
5.2	Conclusion	112
6	Results and Discussion	113
6.1	Introduction	113
6.2	Summary of the Overall Results	113
6.2.1	Investigating Lightweight EfficientNet Models for Precise Semantic Segmentation of CT Images	114
6.2.1.1	Key Observations	115
6.2.2	Comprehensive Analysis of the Results in the Paper: FPN-SE-ResNet Model for Accurate Diagnosis of Kidney Tumors Using CT Images	116
6.2.2.1	Key Observations	117
6.2.3	Comprehensive Analysis of the Results in the Paper: Accurate Kidney Tumor Semantic Segmentation Using Deep Learning Techniques on CT Images	118
6.3	Conclusion	120
6.4	Contribution to Knowledge	121
6.5	Future work	121

List of Figures

1.1	Diagram showing human kidney anatomy and kidney cancer begins in your kidney cells [13].	2
1.2	Examples of kidney tumors: (A) seresnet50 image normal and (B) seresnet50 image with tumor [29].	4

List of Tables

6.1	Performance Metrics and Trends.	115
6.2	Summary Table of Best Models.	116
6.3	Performance Evaluation of SE-ResNet Models Using IoU Metrics.	116
6.4	Performance of DenseNet with LinkNet Using Mean IoU and F1-score. . . .	118
6.5	Performance Comparison of Backbones with Different Architectures. . . .	119

Abbreviations

The following abbreviations are used in this manuscript:

Notations	Description
ANN	Artificial Neural Networks

Chapter 1

Introduction

1.1 Introduction

According to the World Health Organization's report, cancer is regarded as the second most prevalent cause of mortality worldwide [1]. And kidney tumor is the 9th most common cancer among men and the 14th most common among women [2] [3]. The human body has two kidneys on either side of the spine, lower right and left sides, with the right kidney slightly lower due to the liver and the left kidney being more medial in the abdominal cavity [4]. The kidneys are reddish-brown, bean-shaped organs [5]. The kidneys, consisting of millions of nephrons in the renal cortex, maintain a healthy balance of fluids, electrolytes, and organic solutes by filtering, reabsorbing, and secreting hydrogen ions and wastes [6]. Kidney disease affects the kidneys' structure and function, with common types including Chronic Kidney Disease (CKD), Acute Kidney Injury (AKI), Kidney Stones, Polycystic Kidney Disease (PKD), and kidney tumor Figure 1 [7]. Kidney tumors are becoming more common, and the disease develops long without causing symptoms [8]. Renal Cell Carcinoma (RCC) stands as the most commonly diagnosed genitourinary cancer [9] [10] [11]. The exact cause of kidney cancer remains unknown, but several risk factors have been identified. These include smoking, obesity, poor diet, heavy alcohol consumption, family history of hypertension, exposure to radiation and chlorinated chemicals, and heredity, all of which can increase the likelihood of developing the disease [12]. Early detection and prompt treatment are crucial in significantly improving the chances of curing affected individuals [9]. Accurate identification of kidneys and tumors from medical images, like Computed Tomography CT scans, is vital for appropriate treatment. Kidney cancer develops from kidney cells and can spread slowly or precipitously. Manual segmentation of ROI can be tedious and time-consuming, making precise auto-segmentation tools essential [9]. Radiologists analyze (CT) scan data to identify subtle changes in anatomical structures, aiding in diagnosing specific conditions. Researchers seek an automated system to assist radiologists in diagnosis and treatment planning.

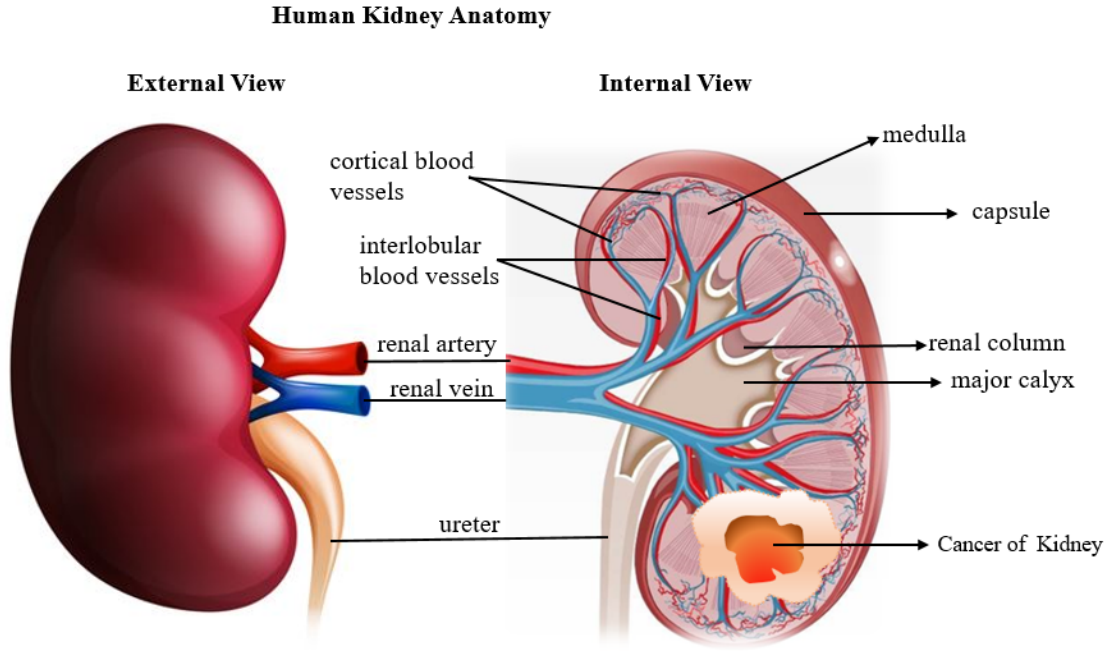


Figure 1.1: Diagram showing human kidney anatomy and kidney cancer begins in your kidney cells [13].

The current diagnostic work-up, utilizing standard imaging instruments like ultrasound, CT and MRI struggles to differentiate RCC from benign lesions in the preoperative setting accurately [14]. Renal lesion biopsy is a highly accurate diagnostic tool for renal cancer, but it is an invasive procedure that poses challenges in specific localizations [14] [15]. A new area of interest in medical research is Computer -aided diagnosis (CAD), which employs artificial intelligence (AI) and its subset machine learning (ML), as well as the state-of-the-art ML approach to deep learning (DL) [16]. Computer-aided diagnosis systems are advanced expert and intelligence systems that combine computer science and medicine. They can simulate the diagnostic decision-making process of a human expert, serving as expert systems. CAD systems can analyze clinical data and derive new knowledge, improving their performance and enhancing current diagnostic criteria. This intelligent CAD system can learn from various data sets, making it intelligent in this regard [17]. Intelligent CAD systems utilize AI, data mining, and machine learning to analyze complex clinical data, formulating diagnostic judgments for various medical conditions and diseases, making them highly effective [18].

Researchers developed automated medical image analysis tools in the 1970s and 1990s using low-level pixel processing and mathematical modelling to handle specific tasks [19]. Medical image analysis faces challenges distinguishing organ pixels from background images, leading researchers to propose automatic segmentation methods using machine learning techniques to extract hand-crafted characteristics [20]. In the 1990s, supervised approaches in medical image analysis, such as active shape models, atlas techniques, feature extraction, and statistical classifiers, gained popularity. This shift from human-designed systems towards computer-trainable systems using example data

and feature vectors has led to the development of hand-crafted features [19]. Technological advancements have led to the development of deep learning techniques, which offer more precise and faster results than traditional methods. Semantic image segmentation has shown promising results in medical image analysis applications. Convolutional neural networks (CNNs) are the most successful image-processing model, with multiple layers and a history dating back to the late 1970s [20] [21] [22] [23]. CNNs excel in complex segmentation tasks due to their non-linear feature extraction and encoder-decoder architecture efficiency [24].

Deep learning techniques have proven effective in semantic segmentation, enhancing image understanding and enhancing machine learning divisions [25] [26]. Deep learning techniques for semantic segmentation include region-based and semi-supervised methods. Region-based methods extract free-form regions from images, classify them using region-based classification, and label pixels based on scoring regions [26]. Additional DL categories for semantic segmentation include feature encoder, recurrent neural network, upsampling, deconvolution, enhanced, spatiotemporal, and CRF/MRF methods [25]. Semantic segmentation is a hot research area in medical image processing, especially for abdominal CT scans, with numerous contests encouraging academics to improve segmentation performance [27]. This work investigates state-of-the-art DL methods for CT renal carcinoma segmentation.

1.2 Motivation

The computer vision task of semantic segmentation entails dividing an image into numerous segments, each possessing unique semantic significance [26]. Medical image segmentation, a subset of semantic segmentation, focuses on outlining regions of interest (ROIs) in images like CT and magnetic resonance (MR) scans. The ultimate aim is to separate specific anatomical areas from the surrounding background, setting the stage for in-depth analysis in research and clinical applications, including morphology and volumetric assessments [28]. As an illustration, we present a case involving the segmentation of a kidney tumor from Computed Tomography (CT) imagery, as depicted in Figure 2. Precisely, the segmentation of a kidney tumor can aid in treatment planning and enhance diagnostic capabilities. Nonetheless, the segmentation process is arduous and necessitates the utilization of specialized expertise as a point of reference. As a result, semantic segmentation is considered one of the most critical tasks in medical image computing and is, therefore, widely favored for clinical applications.

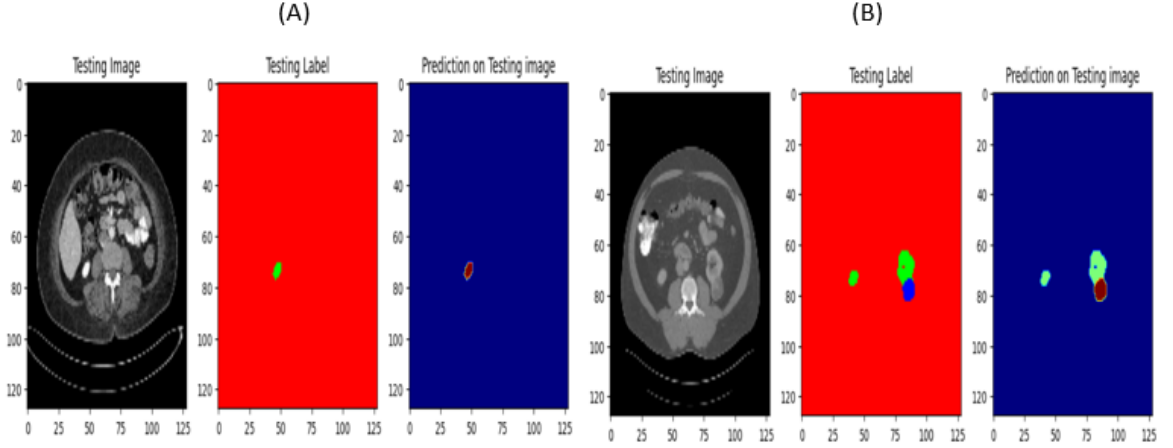


Figure 1.2: Examples of kidney tumors: (A) seresnet50 image normal and (B) seresnet50 image with tumor [29].

This research is subsequently incentivized to conduct extensive investigations into developing a CAD system that is both dependable and effective in detecting and diagnosing renal tumors. To achieve this, sophisticated machine learning techniques, including deep learning methods, are incorporated. Deep learning techniques have demonstrated exceptional efficacy in the analysis of kidney tumors [17]. Their effectiveness is harnessed through their ability to acquire and extract profound features from CT image datasets. Deep learning techniques provide a flexible, fault-tolerant, and efficient solution for kidney tumor diagnosis, outperforming traditional CAD systems. They offer high sensitivity and specificity, enabling precise analysis of kidney tumors, and often outperform traditional approaches in learning distinctive features.

1.3 Problem Statement

Accurately diagnosing kidney tumors in medical imaging remains a significant challenge in the healthcare sector due to several urgent issues. The shortage of specialized medical personnel, variability in image interpretation even among experienced radiologists, and the labor-intensive nature of manually analyzing tumor images hinder the timely detection of kidney cancer. Furthermore, the conventional manual methods are resource-heavy, require substantial time and limit scalability in high-volume healthcare settings.

Advanced computational techniques, such as machine learning and computer vision algorithms, present promising solutions to address these challenges. These technologies could improve the speed, accuracy, and consistency of tumor classification, ultimately enabling earlier interventions and improving patient outcomes.

1. This study focuses on the development and evaluation of deep learning models for kidney tumor segmentation in CT images, addressing key challenges in model efficiency,

feature representation, and diagnostic accuracy. The EfficientNet-U-Net architecture is used to balance computational efficiency with high accuracy, enhancing segmentation performance. High segmentation accuracy is essential for precise tumor measurements, which directly impacts treatment planning. The EfficientNet-U-Net model addresses this challenge by optimizing the design for efficient and accurate kidney tumor segmentation in CT images.

2. The FPN-SE-ResNet model is used to address complex feature representation, diagnostic accuracy, and the need for a model that consistently performs well on diverse tumor characteristics. The model uses a Feature Pyramid Network (FPN) with a Squeeze-and-Excitation (SE) ResNet backbone to enhance the model's ability to capture complex, multi-scale features, improving segmentation accuracy. The FPN-SE-ResNet model leverages global context and localized refinement to enhance diagnostic accuracy in CT-based segmentation.

3. A comparative analysis of three deep learning backbones (EfficientNet, SE-ResNet, and DenseNet) across three segmentation architectures (U-Net, FPN, and LinkNet) is presented to identify the optimal configurations for kidney tumor segmentation. The generalization and reliability of AI models in clinical use are evaluated using the KiTS19 dataset. Quantitative evaluation of segmentation models is also highlighted, with DenseNet121 with FPN achieving the best segmentation scores.

4. In conclusion, this study highlights the importance of a comprehensive, quantitative comparison of deep learning architectures and backbones for kidney tumor segmentation, highlighting optimal combinations that balance accuracy, efficiency and clinical applicability.

1.4 Thesis Objectives

The thesis aims to model a deep learning-based framework for accurate kidney tumour semantic segmentation on CT images. The specific research objectives are:

1. Review state-of-the-art deep learning models specifically used for kidney tumor segmentation on CT images.
2. Identify current gaps, challenges and limitations in existing models, such as segmentation accuracy, computational efficiency and generalizability.
3. Provide a detailed comparison of various architectures to assess which models show the most promise for clinical application and improve diagnostic outcomes.
4. Model a deep learning-based framework for accurate and efficient segmentation of kidney tumors.

1.5 Thesis Contributions

This research makes significant contributions to the field of computer-aided diagnosis (CAD) systems for diagnosing kidney tumors, advancing both the theoretical and practical aspects of kidney tumor segmentation in medical imaging.

Firstly, it introduces EfficientNet U-Net, a model designed to reduce computational demands while maintaining high performance. This innovation enhances model efficiency, making it feasible for real-time applications and suitable for handling large datasets, thereby addressing a critical need in clinical practice.

Secondly, the research achieves improved segmentation accuracy through optimized model configurations. Accurate segmentation is paramount for precise tumor measurements, which are essential for effective treatment planning and prognosis.

The study leverages FPN-SE-ResNet to enhance feature representation. This approach captures complex tumor characteristics across multiple scales, including diverse textures, shapes, and intensities. This advancement addresses the variability seen in kidney tumor presentations, ensuring comprehensive and robust analysis. To support clinical reliability, the developed models prioritize consistency and diagnostic accuracy by ensuring reliable and repeatable segmentation outcomes. This consistency is aimed at reducing errors and improving decision-making in clinical environments.

Furthermore, the research conducts a comparative analysis of architectures and backbones, systematically evaluating the performance of EfficientNet, SE-ResNet, and DenseNet integrated with U-Net, FPN, and LinkNet architectures. This analysis identifies optimal combinations for kidney tumor segmentation, providing valuable insights for future model development. The study also offers quantitative benchmarking, providing clear performance metrics such as Dice scores, IoU, and F1-scores. These benchmarks establish a standard for evaluating and comparing kidney tumor segmentation methods, aiding researchers and practitioners in assessing model efficacy.

Finally, the models are validated using the KiTS19 dataset, ensuring robustness and reliability in diverse clinical applications. This validation emphasizes the models' potential for generalization, supporting their applicability across various medical scenarios. Collectively, these contributions pave the way for the development of accurate, efficient and clinically viable deep learning models tailored to kidney tumor segmentation. This research stands as a critical step toward enhancing the precision and reliability of CAD systems in medical imaging.

1.6 Thesis Outline

The thesis is presented using a submission by published article format. The rest of this thesis is arranged as follows: In Chapter Two, we review the latest advancements in deep learning for kidney tumour segmentation on CT images. Accurate kidney tumor

segmentation is vital for assessing tumor characteristics, planning treatment and tracking disease progression. Manual segmentation is labor-intensive inconsistent and varies across observers underscoring the need for efficient automated solutions.

In chapters Three to five, our survey explores the evolution of deep learning models for this task, including convolutional neural networks (CNNs) and their application in medical imaging. We examine key architectures and backbones used for segmentation and highlight recent models that improve accuracy and computational efficiency. This survey lays the foundation for our study, which investigates advanced architectures and backbone configurations to optimize performance in kidney tumor segmentation, addressing the clinical need for reliable and automated diagnostic tools. Chapter six is a significant milestone as it not only consolidates the findings of the research but also provides a reflective summary and a glimpse into future possibilities inspired by the work completed here.

1.7 Datasets

The dataset utilized for this thesis is primarily from the Kidney Tumor Segmentation Challenge (KiTS19), a publicly available open-source dataset. This dataset was released as part of the Medical Image Computing and Computer-Assisted Intervention (MICCAI) challenge in 2019 and contains annotated CT scans from over 200 kidney cancer patients. The KiTS19 dataset has been extensively used to train, validate, and benchmark all models developed in this thesis, aiding in evaluating and comparing segmentation accuracy across various architectures. All data sources referenced in this thesis have been properly acknowledged.

1.8 Responsible AI

The FPN-SE-ResNet, EfficientNet-based U-Net, and DenseNet with LinkNet models have shown remarkable potential in accurately segmenting kidney tumors from CT images, contributing to advancements in computer-aided diagnosis (CAD) systems. However, to ensure these AI-driven tools are both reliable and ethically applicable in clinical practice, strict adherence to responsible AI principles is crucial.

To achieve high-performance and clinically relevant results, these models were trained on large-scale, well-annotated, and diverse datasets. This approach helps mitigate bias and ensures that the models generalize effectively across different patient demographics, tumor types, and imaging conditions. A diverse dataset is particularly important in medical AI, as it reduces the risk of performance discrepancies across varying populations, thereby improving fairness and reliability in real-world applications.

Furthermore, these AI models prioritize explainability and transparency, enabling radiologists to interpret the model’s decision-making process. By integrating techniques

such as saliency maps, attention mechanisms, and feature visualization, these models provide insights into which image regions influenced their predictions. This interpretability fosters trust among medical professionals, allowing them to verify AI-assisted diagnoses and confidently integrate the technology into clinical workflows.

In addition, the dataset was systematically divided into distinct training, validation, and testing subsets to ensure robust model evaluation. The training set was used to optimize model parameters, the validation set helped fine-tune hyperparameters and prevent overfitting, and the testing set was reserved for assessing real-world performance. This structured data partitioning ensures that the models do not merely memorize patterns from training data but instead develop a genuine ability to generalize to unseen cases.

By implementing these responsible AI practices, the FPN-SE-ResNet, EfficientNet-based U-Net, and DenseNet with LinkNet models are positioned to offer reliable, interpretable, and ethically sound solutions for kidney tumor segmentation, ultimately supporting improved diagnostic accuracy and patient care in clinical settings.

Chapter 2

Literature Review and Related Works : Kidney Tumor Semantic Segmentation Using Deep Learning: A Survey of State-of-the-Art

2.1 Introduction

This section presents a research paper that conducted a comparative analysis of recent studies on kidney tumor detection. The review paper thoroughly examined advanced deep learning techniques, including an in-depth discussion of commonly used convolutional neural network architectures for kidney tumor analysis, along with a critical evaluation of the performance of specific deep learning models and the standard metrics for assessing their effectiveness. Additionally, the paper explored vital factors that could enhance the performance of kidney tumor detection systems in the future.

Article

Kidney Tumor Semantic Segmentation Using Deep Learning: A Survey of State-of-the-Art

Abubaker Abdelrahman  and Serestina Viriri * 

School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal,
Durban 4000, South Africa; 220106722@stu.ukzn.ac.za

* Correspondence: viriris@ukzn.ac.za

Abstract: Cure rates for kidney cancer vary according to stage and grade; hence, accurate diagnostic procedures for early detection and diagnosis are crucial. Some difficulties with manual segmentation have necessitated the use of deep learning models to assist clinicians in effectively recognizing and segmenting tumors. Deep learning (DL), particularly convolutional neural networks, has produced outstanding success in classifying and segmenting images. Simultaneously, researchers in the field of medical image segmentation employ DL approaches to solve problems such as tumor segmentation, cell segmentation, and organ segmentation. Segmentation of tumors semantically is critical in radiation and therapeutic practice. This article discusses current advances in kidney tumor segmentation systems based on DL. We discuss the various types of medical images and segmentation techniques and the assessment criteria for segmentation outcomes in kidney tumor segmentation, highlighting their building blocks and various strategies.

Keywords: kidney tumor segmentation; deep learning; computerized tomography imaging; survey



Citation: Abdelrahman, A.; Viriri, S. Kidney Tumor Semantic Segmentation Using Deep Learning: A Survey of State-of-the-Art. *J. Imaging* **2022**, *8*, 55. <https://doi.org/10.3390/jimaging8030055>

Academic Editor: Caroline Petitjean

Received: 19 November 2021

Accepted: 10 February 2022

Published: 25 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The kidney is an organ of a vital role that keeps the body fluid and solute balance checked by excreting and filtering waste products. It also secretes many hormones and helps control blood pressure. The human kidneys are depicted in (Figure 1) [1]. Kidney cancer is one of the top 10 malignancies in men and women alike. The probability of having kidney cancer throughout one's lifetime is around 1 in 75 (1.34%) [2]. Renal cancer (RC) is an acute urological disease that affects over 400,000 individuals each year [3,4]. According to the Global Cancer Observatory (GCO), more than 175,000 deaths are due to this disease [5,6]. Renal cell carcinoma (RCC) has the third-highest disease rate after prostate cancer and bladder cancer [7]. It is estimated that RCC is the seventh most frequent cancer in men and the ninth most common cancer in women in the United States, with 48,780 new cases diagnosed and 27,300 new instances of RCC-related death [8]. On radiography, distinguishing between benign kidney tumors and malignant renal cell carcinoma can be challenging [9]. However, the majority of kidney tumors turn out to be cancerous [10]. Renal cell carcinoma (RCC) represents the vast majority of these tumors [11,12]. Clear cell RCC is the most common subtype of renal RCC [13], accounting for approximately 80–90% of all kidney cancers. Overall, the worldwide incidence rate has increased by 2% per year during the last two decades [8]. Kidney tumors are becoming more common, and the disease develops for a long time without causing symptoms. Coincidence may be the reason for detecting more than half of the cases of renal cell carcinoma [14,15]. Importantly, the cause of kidney cancer has not been determined. However, many risk factors, including smoking, obesity, poor diet, substantial alcohol consumption, family history of hypertension, exposure to radiation and chlorinated chemicals, and heredity, are risk factors for getting sick [16]. Radical nephrectomy (RN) and partial nephrectomy (PN) are the current therapeutic options for kidney tumors. The tumor and kidney are removed in RN, but in PN, only the malignancy is removed [17].

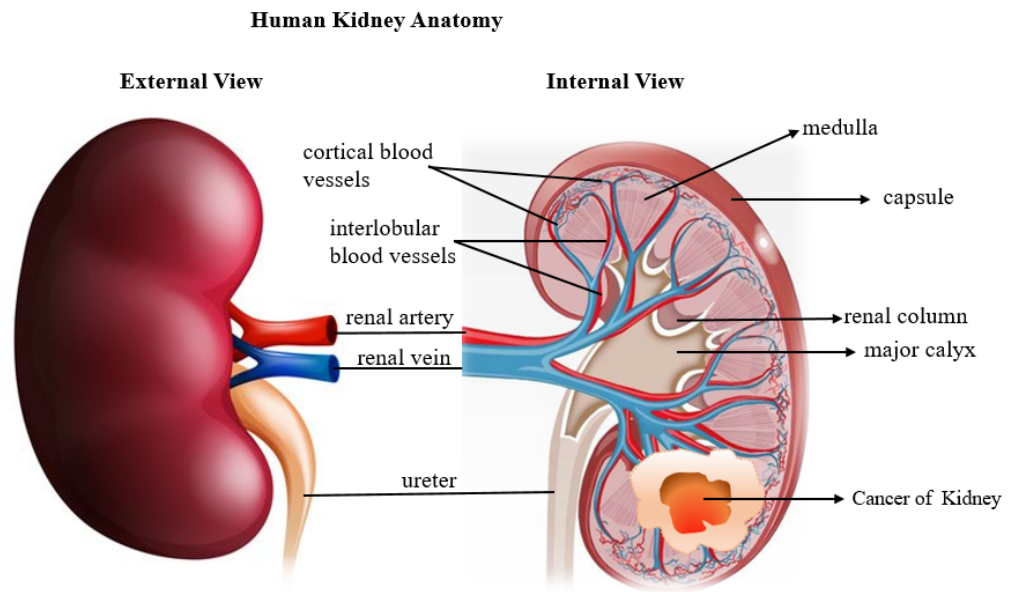


Figure 1. Diagram showing human kidney anatomy and Renal cell carcinoma developed inside the kidney.

Image processing is a widely utilized technology in a range of industries, including robotics, biometrics, security and surveillance, remote sensing, and medical imaging. An image processing task's overall applicability and performance are heavily dependent on the quality of the test image [18]. Medical imaging techniques are divided into several types: ultrasound sonography (US), computed tomography (CT), and magnetic resonance imaging (MRI). Medical images (MI) have excellent homogeneity, make it challenging to identify regions of interest and patterns and blurring the boundaries between organs and other areas. Radiologists favor CT imaging over other imaging modalities because it produces high-resolution images with good anatomical features. In addition, it gives images with excellent contrast and exceptional spatial resolution. Therefore, CT imaging is an essential tool for diagnosing any disease affecting the kidneys [1]. It is frequently utilized in clinics for therapy planning and segmentation of kidney tumors [19]. In addition to that, some CT results can be utilized to classify benign cancers (Figure 2) [20]. The most common method of tumor delineation is by hand. An expert radiologist will carefully examine the patients scanned medical photographs, segmenting all damaged areas. Manual segmentation is time-consuming. It also has a lot of intra- and inter-rater variability [21]. CT technologies can significantly enhance our ability to detect and monitor diseases and patients, and this may improve patient care and facilitate the evaluation of treatment strategies.

Image segmentation is the process of splitting an image into several segments in order to transform it into a more meaningful and easy-to-analyze representation [22]. The process of image segmentation may be conceived of in two steps: identification and delineation. Identification is the process of identifying the location of an object in an image and differentiating it from everything else in the image. Segmentation involves delineating the boundaries of the region of interest for further analysis [1]. There are several methods for segmenting images: Manual Segmentation, Semi-Automatic, Automatic Segmentation, and Semantic Segmentation. Semantic segmentation is crucial for image analysis tasks and plays a significant part in image interpretation. Image categorization, object recognition, and border localization are all required for semantic segmentation [23]. Semantic segmentation has several applications in computer vision and artificial intelligence-assisted autonomous driving, and medical imaging analysis [24–26].

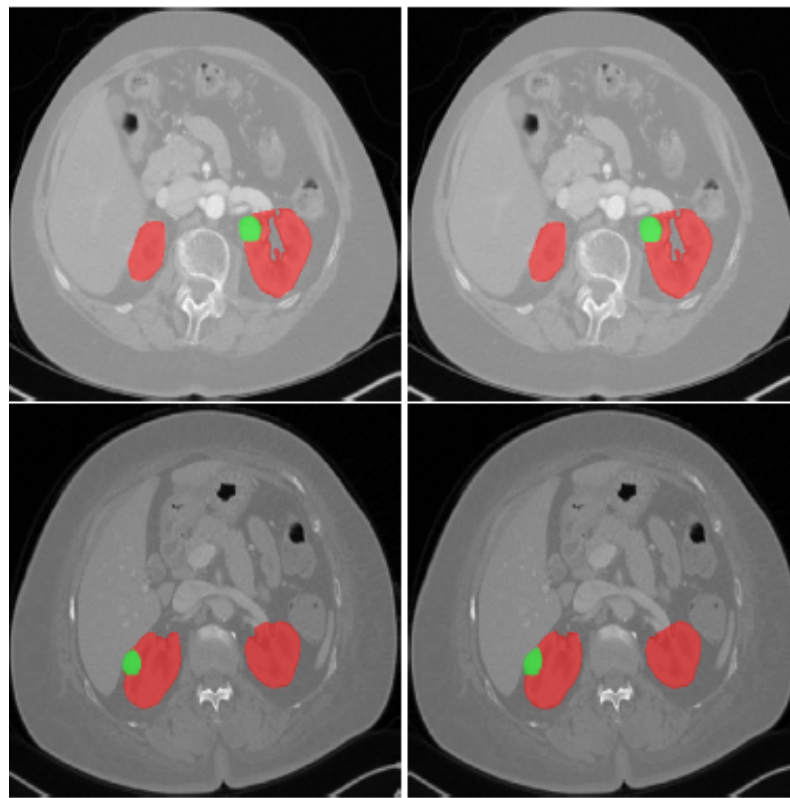


Figure 2. CT scans showing: An axial slice of two patients' 3D CT scans from the KiTS19 dataset; the red tint denotes the kidneys, whereas the green color indicates the tumor site [5].

Since scanning and loading (MI) onto a computer became practical, researchers have created automated analysis tools. MI analysis was done between the 1970s and 1990s by combining low-level pixel processing (edge and line detector filters, region expansion) with mathematical modeling (fitting lines, circles, and ellipses) to create compound rule-based systems that handled specific tasks [27]. One of the most challenging problems in medical image analysis (MIA) using traditional approaches such as edge detection filters and mathematical algorithms is distinguishing the pixels of organs or lesions from background medical images of CT or MRI scans to give vital information on the shapes and sizes of these organs. Therefore, researchers have suggested numerous automatic segmentation methods to extract the hand-crafted characteristics, such as machine learning techniques [28]. Around the 1990s, supervised approaches involving training data to develop a system became more prevalent in medical image analysis. Active shape models (for segmentation), atlas techniques (in which atlases are fitted to fresh data extracted from the training data), feature extraction, and statistical classifiers are just a few examples (for computer-aided detection and diagnosis). This pattern recognition or machine learning technique is still frequently utilized, and it is the basis for a large number of commercially accessible medical image analysis products. As a result, we have seen a shift away from human-designed systems toward systems that computers train using example data and extract feature vectors. In the high-dimensional feature space, computer algorithms find the best decision boundary. The extraction of discriminant characteristics from images is a critical step in the construction of such systems. Humans still carry out this procedure, and as a result, one speaks of systems with hand-crafted features [27]. As a result of these technological advancements, deep learning techniques began to exhibit their significant capabilities in image processing applications [28]. DL is a type of machine learning that enables more precise and quicker results than traditional machine learning techniques [29]. Recently, DL techniques for semantic image segmentation have shown promising results in a variety of medical image analysis applications [30–32]. Convolutional neural networks (CNNs) are the most successful form of image processing model to date. CNN have

multiple layers [27]. CNN has been under development since the late 1970s by Fukushima. Moreover, they were used to analyze medical images in 1995 [33]. They witnessed their first successful real-world application in 1998 for hand-written digit recognition. In the December 2012 ImageNet competition, Alex-Net, the planned CNN, won the competition by a huge majority. Using analogous but deeper designs, more work was done in later years. Deep convolutional networks have emerged as the preferred method for computer vision [27]. Due to their powerful non-linear feature extraction capabilities and the efficiency of encoder–decoder architectures, CNNs have been used for complex segmentation tasks [5]. In computer vision tasks, CNN architectures have already surpassed classical methods [12], particularly in the segmentation of CT scans [34]. On the other hand, fully convolutional network (FCN) architectures are a notably powerful end-to-end training segmentation solution. This architecture is state-of-the-art across several domains and can create raw-scale, pixel-level labels on the images [12]. Other subsequent efforts, such as the feature pyramid networks (FPN), primarily utilized in object recognition, have used FCN as a starting point for deeper and more sophisticated segmentation structures [35]. Pyramid scene parsing networks (PSPNets) are used to analyze scenes [36]. For object instance segmentation, Mask R-CNN [37] is used.

Semantic segmentation using DL techniques: DL, a proliferating new machine learning division, has proven their effectiveness in semantic segmentation. Deep learning techniques play an important role in easing image understanding [23,38]. Deep learning techniques for semantic segmentation have been divided into region-based, FCN, and semi-supervised methods. Region-based methods adopt the pipeline method by first extracting free-form regions from input images, then classifying them using region-based classification; finally, they label pixels according to the scoring regions [38]. FCN-based methods, in contrast to region-based methods, do not extract the region proposal. They learn a mapping from pixel to pixel, making arbitrary-sized images [38]. Regarding semi-supervised methods, usually, semantic methods depend on many images that require a long time to annotate the masks. Therefore, some semi-supervised methods have been suggested to utilize the annotation process [38]. In addition to the methods mentioned, more DL categories for semantic segmentation have been proposed, including feature encoder-based methods, recurrent neural network-based methods, upsampling/deconvolution-based methods, increased resolution of feature-based methods, enhancement of feature-based methods, spatiotemporal-based methods, and methods using CRF/MRF [23].

Semantic segmentation is a hot research area in medical image processing. This is particularly true for abdominal CT scans, where many contests push academics to continue developing approaches for improving segmentation performance [39]. Even though medical segmentation is becoming increasingly prevalent, there are few ways for segmenting kidney and renal tumors in the literature [40]. To our knowledge, only a few review articles have examined kidney segmentation strategies. Nonetheless, numerous articles have been published on the subject of kidney segmentation. Additionally, various deep learning-based algorithms have been developed for segmenting images of other diseases, such as skin lesions, dental imaging, and eye images. As a result, the research barriers for adopting and applying these approaches to the kidney images have encouraged study on kidney segmentation and detection. This study look at cutting-edge DL techniques for segmenting CT kidney cancers. Additionally, the article highlights distinct challenges and possible solutions for medical image interpretation.

2. Related Work

In recent years, we have seen tremendous advancement in development in many fields as processing power have increased. Among the technologies that have succeeded and evolved are those that are based on deep learning, most notably CNNs, which have been successful in medical image processing. It has risen to the top of the list for various medical image analysis applications, including medical image segmentation detection, abnormality classification, and medical image retrieval. Many interesting techniques have

appeared in the field of medical image segmentation using DL [41]. Moreover, the problem of kidney image segmentation has attracted research earlier. Some approaches, such as atlas-based methods, deformable models, graph cuts technique, and region growing, have been developed and used. In [42] the authors describe the shapes of the local objects via contour templates, which are used to capture specific properties of different organs. The authors in [43] used 2D CNN for the segmentation of computerized tomography (CT) images. At present, end-to-end segmentation and cascade segmentation are the two most commonly used deep learning segmentation strategies in image segmentation. Cascaded segmentation refers to multistage processing to achieve stepwise segmentation, whereas end-to-end segmentation employs only one model to execute the segmentation operation directly. On the other hand, end-to-end segmentation better prevents the buildup of mistakes in multistage segmentation and streamlines the procedure. However, a single model with great integration reduces its flexibility, operability, and interoperability. Simultaneously, a single model may require additional training data in order to obtain better outcomes. As a result, many researchers still use cascade approaches in medical image segmentation [41].

2.1. One-Stage Methods

One-stage methods [39,44–56] are designed to predict the multi-class results directly from whole images. Myronenko et al. [44], from arterial phase abdominal 3D CT images, presented an end-to-end boundary aware fully CNN for accurate kidney and kidney tumor semantic segmentation. Efremova et al. [45] combined U-Net and LinkNet-34 with ImageNet-trained ResNet-34 to decrease the convergence time and overfitting. Their model has shown success in a wide range of computer vision applications, including medical image processing. Guo et al. [46] proposed an automatic segmentation model called RAU-Net. Their model has been developed for renal tumor segmentation. With some utilization of the cross-entropy function, the model can identify positive samples. However, their generalizability is lacking. Isensee et al. [39] designed a U-Net model that performs well with the KiTS2019 dataset and is able to learn segmentation tasks from reference data. To achieve a regularizing effect, you can either lower the number of layers in the typical U-Net or increase the number of residual blocks. Causey et al. [47] proposed a deep learning model (Arkansas AI-Campus) that collects U-Net models produced after several model variants were tested. Their model performs consistently on both the local test dataset and the final competition independent test dataset; it takes place in the top five of the KiTS19 Competition among US teams. Nazari et al. [48] developed a DL technique to detect the borders of organs with high accuracy using computed tomography images. They used the obtained inches to calculate dosimetry using cautery as the source organ. Yasmeen et al. [49] demonstrate a deep neural network cascaded for semantic segmentation of kidneys and surrounding anatomy. Ruan et al. [50] much of the work focuses on the feature map at the network's bottom, which enhances network performance by extracting and fusing multi-scale information. Yu et al. [51] it is recommended that non-squared patches with varying aspect ratios be used to train segmentation networks in order to integrate more global contexts in local details. According to Pang et al. [52], automatic image segmentation is a frequent application case in machine learning that has gotten much attention in recent years. Tumor segmentation in computed tomography (CT) images is a popular application. Shen et al. [53], to cope with kidney and tumor segmentation problems, suggested the COTRNet. COTRNet uses a transformer to capture long-range dependencies for accurate tumor segmentation, inspired by the DETR, which used one to represent global characteristics. Yang et al. [54] suggested a 3D fully convolutional network with a pyramid pooling module intended specifically for segmenting kidney and renal pathologies. Experiments and comparisons with other methods show that their method performs very well, with an average dice coefficient of 0.931 for kidney segmentation and 0.802 for tumor segmentation. To make an attempt at resolving the class imbalance, Heo et al. [55] presented a one-stage semantic segmentation model based on 3D abdominal CT imaging for the KiTS21 Challenge. The model was constructed using U-Net and the sum

of the Focal and Dice Losses. To improve performance, Christina et al. [56] benefited from a strategic sampling approach based on clinical data. Using random sampling, a baseline 3D U-Net was trained. The clinical features most strongly related to segmentation success were determined using LASSO regression and then included into a mindful sampling method, maximizing the influence of the identified clinical characteristics.

2.2. Two-Stage Methods

The goal of the two-stage approach is to overcome the problem of foreground/background imbalance. Those approaches begin by detecting the volume of interest (VOIs) and then segmenting the target organs from the VOIs (VOIs) [57]. Regarding two-stage methods [12,41,58–69], Cruz et al. [58] developed a method that uses deep convolutional neural networks with image processing techniques to delimit the kidneys in CT images, where they achieved up to 93.03% accuracy, so further improvements are required. Zhang et al. [59] studied a cascaded two-stage framework using a 3D fully convolutional network (FCN) for kidney and tumor segmentation. Their method locates the kidney and cuts off the irrelevant background. Hou et al. [60] offered a triple-stage self-guided network for the kidney tumor segmentation challenge. From down-sampled CT images, the low-resolution net can approximately find the volume of interest (VOI). Still, the full-resolution net and tumor refine net can extract the right kidney and tumor borders inside VOI from full-resolution CT images. Their model squanders computational resources while training numerous networks. Hatamizadeh et al. [61] enhanced the edge representations in learned feature maps with their module that can be combined with any generic encoder–decoder architecture; the core idea of their model is to add an extra task of learning edges to the original network. Some researchers attempt to push their networks to learn specific characteristics in an ambiguous and multitasking manner, which is unacceptably inefficient. Zhao et al. [12] developed a U-Net-based model called MSS U-Net, a multi-scale supervised 3D U-Net for segmenting kidneys and kidney cancers from CT scans. They combined deep supervision with exponential and logarithmic loss to improve the efficiency of 3D U-Net training. Santini et al. [62] combined Res-Net with Res-U-Net architectures in a multi-stage DL approach called EMS-DLA that has been used for kidney tumor segmentation. The results are promising, and they might be improved if an enhanced understanding of benign cysts is factored in. Xie et al. [41] presented a cascaded SE-ResNeXT U-Net. Chen et al. [63] presented a technique for segmenting kidney, tumor, and cyst in abdomen enhanced CT images based on a multi-stage stepwise refinement strategy. In network training, a 2.5D technique is utilized for data input to maintain certain contextual semantic information while reducing memory strain. There are certain points in this report that need to be further looked at. For smaller kidneys, tumors, and cyst segmentation, the network and procedures can be enhanced. Wei et al. [64], to differentiate kidney cancers, employed two-phase models, which are cascaded network structures. In a tumor, they achieved 0.75. He et al. [65], for kidney segmentation, suggested a novel two-stage cascade and multi-decoding approach. They used U-Net to locate and extract the kidney area and then MSD-Net for final segmentation. Yi Lv et al. [66] offered a three-step automated kidney tumor and cyst segmentation approach based on 3D U-Net. According to the findings, the average dice for kidneys, tumors, and cysts is around 0.93, 0.57, and 0.73, respectively. The accuracy of tumors and cysts, on the other hand, is not sufficient. Li et al. [67] presented a two-stage cascaded segmentation technique for the kidney, renal tumor, and renal cyst. This was accomplished by embedding a Residual 3D U-Net architecture into each level of the cascaded process. The suggested approach demonstrated good segmentation outcomes for the kidney and tumor. Because the border between the tumor and the kidney is ambiguous, the segmentation of the kidney and tumor is complicated. Xiao et al. [68] suggested a two-stage detection and segmentation architecture for autonomously segmenting kidneys, cysts, and tumor based on the KiTS21 benchmark. The ResUnet 3D was used as the foundation. The two-stage architecture produced a mean dice of 0.6543 for kidney and messes, 0.6543 for kidney messes and tumor, and 0.4658 for the mean surface dice. Wen et al. [69]

They presented a unique segmentation network named SeResUNet to segment the kidney and tumor. Here, one must choose an encoder–decoder architecture such as U-Net and utilize ResNet to deepen the encoder’s network and, simultaneously, to minimize severe network degradation and accelerate convergence.

2.3. Hybrid Models

Tow introduced hybrid models in [70–72]. Abdul Qayyuma et al. [70] designed a hybrid 3D residual network (RN) with a squeeze-and-excitation (‘SE’) block to acquire spatial information by utilizing cSE blocks. The reweighting function in a “three-D RN” is used. Their network has been tested on various datasets and performs well in medical image segmentation, especially in volumetric segmentation of the kidney, liver, and related malignancies. Cheng et al. [71] enhanced 3D SEAU-Net to develop a multi-class segmentation architecture to improve the performance. Their model aggregates residual network, dilated convolution, squeeze-and-excitation network, and attention mechanism. The multi-class segmentation job is decomposed into two more easy binary segmentations. Cruz et al. [72] provide an effective approach for segmenting kidney cancers in computed tomography. Thus, post-processing was applied on the DeepLabv3+ 2.5D model with DPN-131 encoder. Additionally, image processing methods such as normalization, proportional distribution of training and validation datasets, and DART were employed. Finally, the findings were achieved by integrating preprocessing, segmentation of the kidney tumor, and postprocessing.

3. Overview of Deep Learning (DL) Models

This section discusses the DL concepts, techniques, and architectures of DL algorithms for kidney and renal tumor segmentation that we discovered when reviewing the medical image analysis articles examined in this work, as summarized in Figure 3.

DL is an artificial intelligence (AI) function concerned with decision-making algorithms inspired by the structure and function of the human brain, referred to as “artificial neural networks”. It is a branch of machine learning in artificial intelligence that utilizes networks capable of unsupervised learning from unstructured or unlabeled input. They are also referred to as deep neural networks or deep neural learning. In a variety of medical image analysis applications, semantic image segmentation algorithms based on deep learning have demonstrated promising results. Deep CNNs have been state-of-the-art for many image classification and segmentation applications. CNNs have been used for complex segmentation problems due to their superior non-linear feature extraction capabilities and the efficacy of their encoder–decoder architectures [73].

3.1. Neural Networks

Neural networks are a sort of learning algorithm that serves as the foundation for the majority of DL techniques [21,27]. As seen in Figure 4, neural networks are nothing more than a set of arithmetic operations that convert an input (x) into an output (y). Weights and linear operators multiply the previous value of connected input and output to the neuron. The function in the hidden layer can be anything. In the simplest example, with a single neuron in the hidden layer (A), the input value is multiplied by the first weight, and the resulting value $x \times \text{weight1}$ is then passed to the neuron’s function. The result is calculated by multiplying the output of that function by weight2. When a neural network is trained, it is fed inputs and the outcome is computed using mathematics. The output value is compared to the known real value of y ; the weights are slightly modified to get the output value closer to the known actual value of y . A straightforward illustration of this is presented in (B), where weight1 equals 2, function equals $2 \times x$, and weight2 equals 2 [74].

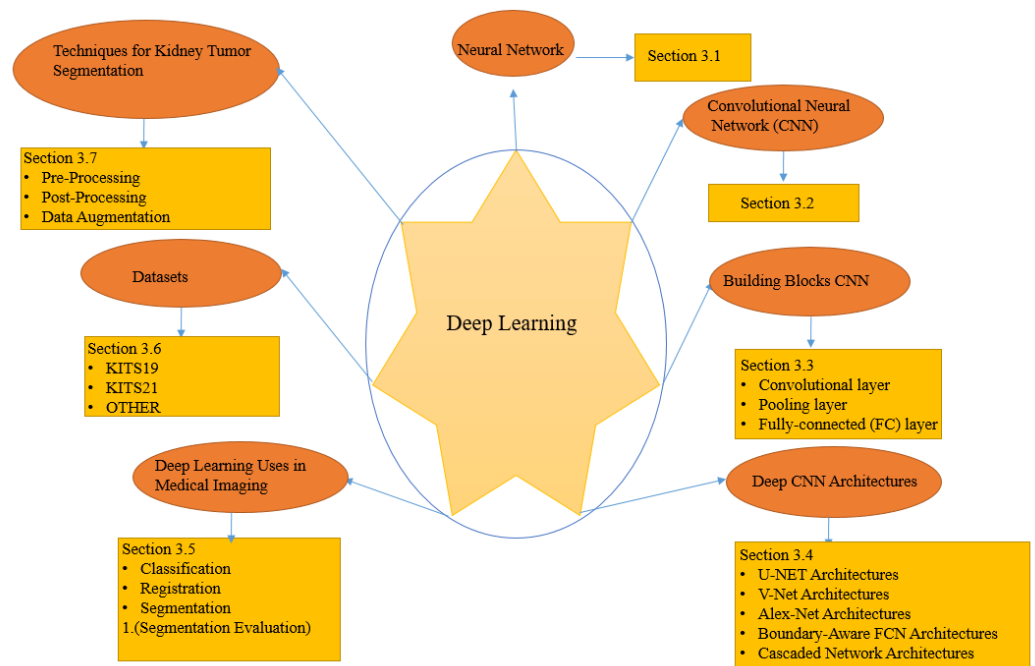


Figure 3. Components architectures, and strategies for deep learning algorithms for segmenting kidney tumors.

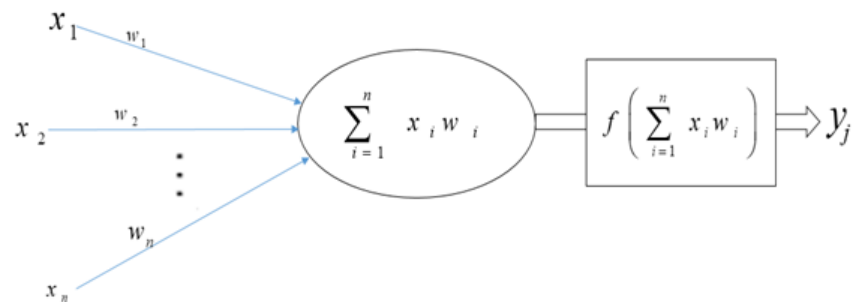


Figure 4. The building block of deep neural networks is an artificial neuron or node.

Each input x_i has a corresponding weight w_i in an artificial neuron or node. The total of all weighted inputs (x_i, w_i) is then fed into a nonlinear activation function (f), which translates the preactivation level of the neuron to an output (y_j). The bias terms have been deleted for simplicity. The result (y_j) is then used as an input for the following layer's node. Numerous activation functions are available, each with a somewhat different mapping of a preactivation level to an output value. The rectifier function is the most often activated function (neurons that employ it are referred to as “rectified linear units” (ReLU)), followed by the hyperbolic tangent function, the sigmoid function, and the softmax function. As seen in Figure 5, the latter is frequently utilized in the output layer to compute the likelihood of multiclass labels.

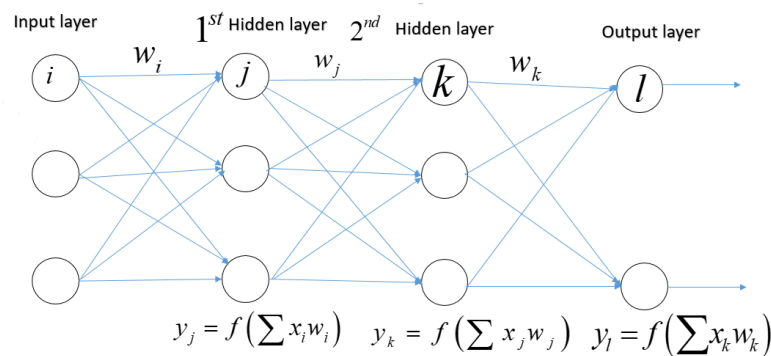


Figure 5. Artificial feedforward multilayer neural network.

A feedforward multilayer neural network with two classes (also known as a multilayer perceptron). Each node in one layer is linked to every neuron in the subsequent layer (a fully connected network). The weighted total of the inputs for each neuron j in the first hidden layer is multiplied by a nonlinear function. The output (y_j) of this transformation is utilized as the input for the second hidden layer. The information is transferred to the output layer through the network.

3.2. Convolutional Neural Network (CNN)

A CNN is a kind of neural network. It was first released in 1980 [75]. CNNs have established themselves as the de facto technique for handling a wide variety of challenging computer vision issues in recent years [76]. CNN is one of the most recent DL algorithms in image recognition. CNN is inspired by the multi-layered structure of the visual cortex in the human brain and has demonstrated outstanding performance in a variety of very complicated application scenarios [76,77]. The traditional machine learning technique to image identification consists of two distinct processes. Using several techniques, such as HOG [78], SURF [79], or HOU [80], the initial stage, dubbed feature engineering, attempts to extract meaningful data representations from the raw image data. In the second stage, referred to as classification, a machine learning algorithm attempts to discover a pattern that links previously created data representations to a target variable. The algorithm is only capable of learning these patterns if they have already been retrieved via feature engineering. Manual extraction of appropriate data representations, in particular, frequently results in unsatisfactory categorization results [76]. The combination of these two phases is the essential distinction between convolutional neural networks and standard machine learning algorithms for computer vision [76]. CNN are feedforward neural networks intended to analyze images and are physiologically inspired by the visual cortex [81]. Numerous convolutional layers are used in common CNN topologies. Each image is processed in three dimensions using a three-dimensional array. By applying several tiny filter kernels to the image array, the convolutional layers convert the original input to feature maps [76], as seen in Figure 6. The filter matrices are applied to the entire image, preserving spatial information. Following that, these feature maps are processed via a nonlinearity function such as ReLU, a batch-normalization layer, a convolutional layer, and a pooling layer. CNN are capable of automatically extracting useful feature representations with fully connected layers from raw images and optimizing them to represent specific target classes by combining several convolutional, activation, batch normalization, and pooling layers [76]. In comparison to alternative models, CNN has recently become the de facto model for medical image segmentation due to its record-breaking performance in conventional computer vision tasks as well as medical image analysis [21]. CNN models may learn spatial hierarchies of features within data, e.g., the first layer will learn tiny local patterns, such as edges, while the second convolutional layer may learn bigger patterns constructed from the first layer characteristics, and so forth. They are more suited to image analysis jobs because of this capacity. Furthermore, convolutional layers' units share weights, decreasing the

number of parameters to train and improve the network's efficiency [21]. CNN has recently become the de facto paradigm for kidney tumor segmentation because of its superior performance in traditional computer vision and medical image analysis when compared to alternative models. CNN models can learn to create spatial hierarchies of features included inside data.

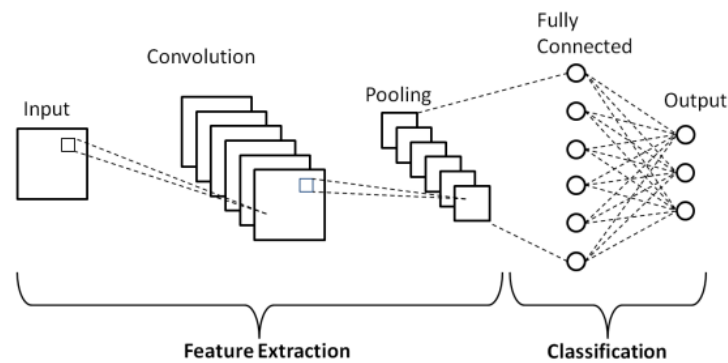


Figure 6. Schematic diagram of a basic convolutional neural network (CNN) architecture [19].

A CNN's architecture is hierarchical. Starting with the input signal, each succeeding layer x_j is calculated as follows [82]:

$$x_j = PW_j X_j - 1$$

Here W_j is a linear operator, and ρ is a non-linearity. Typically, in a CNN, W_j is a convolution, and ρ is a rectifier $\max(x, 0)$ or sigmoid $\frac{1}{1+\exp(-x)}$. It is easier to think of the Operator W_j as a stack of convolutional filters. Thus, the layers are filter maps and each the layer can be expressed as the sum of the previous layer's convolutions [82].

$$x_j(u, k_j) = \rho\left(\sum_k (x_j - 1(\cdot, k) * W_j, k_j(\cdot, k))(u)\right)$$

Here is the discrete convolution operator:

$$(f * g)(x) = \sum_{u=-\infty}^{\infty} f(u)g(x - u)$$

A CNN's optimization issue is substantially non-convex. Thus, the weights W_j are often taught by stochastic gradient descent, with the gradients computed using the back-propagation process [82].

3.3. Building Blocks CNN

A convolutional network's initial layer is the convolutional layer. Convolutional layers can be followed by additional or pooling levels, but the wholly linked layer is the final layer. The CNN becomes more sophisticated with each layer, recognizing larger areas of the image. Earlier layers concentrate on essential elements like colors and borders. As the image data goes through the CNN layers, and it detects more significant components or forms of the item, eventually identifying the desired object. They have three main types of layers, which are:

- Convolutional layer.
- Pooling layer.
- Fully-connected (FC) layer.

3.3.1. Convolutional Layer

This layer will determine the output of neurons connected to local regions of the input through the calculation of the scalar product between their weights and the region

connected to the input volume. This layer is the central constituent of a CNN, and it is here that the majority of the processing takes place. It requires input data, a filter, and a feature map, among other things. This layer consists of a set of learnable filters or kernels (the typical size is usually 3×3 or $3 \times 3 \times 3$, depending on whether the input is a two-dimensional (2D) or three-dimensional (3D) image, respectively) [21].

3.3.2. Pooling Layer

A pooling layer is typically used in conjunction with a convolutional layer or a collection of convolutional layers. The objective is to minimize the size of the feature maps while retaining critical features. A sliding window pooling method is conducted to a rectangular neighborhood. For instance, max pooling is used to maximize the size of a rectangle neighborhood. Additionally, average and weighted average pooling are prominent pooling methods [21]. Similar to the first layer, the pooling procedure sweeps a filter across the whole input, except this filter does not include any weights. Rather than that, the kernel populates the output array with values from the receptive field using an aggregation function. Pooling may be classified into two types: Max pooling, Average pooling. Pooling is primarily concerned with down sampling in order to lessen the complexity of subsequent layers. In the sphere of image processing, this is analogous to decreasing the resolution. The number of filters does not change as a result of pooling. Max-pooling is one of the most often used pooling techniques. It divides the image into rectangular sub-regions and returns just the largest value contained inside each sub-region. One of the most often utilized sizes in max-pooling is 2×2 [82].

3.3.3. Fully Connected (FC) Layer

The moniker “fully connected layer” is self-explanatory. As previously stated, with partially linked layers, the pixel values of the input image are not directly connected to the output layer. By contrast, each node in the output layer connects directly to the previous layer in the eventually linked layer. This layer performs classification tasks based on the characteristics collected by the preceding layers and their various filters. While convolutional and pooling layers often employ ReLU functions to categorize inputs, (CL) are used to extract features. The characteristics they generate are subsequently classified by the fully connected (FC) layers. As seen in Figure 6, each unit in the FC layer is connected to every unit in the preceding layer. Typically, the last layer is a softmax classifier that generates a probability vector map across the various classes. Prior to passing the features to an FC layer, they are all transformed to a one-dimensional feature vector. This results in the loss of spatial information included in image data. A disadvantage of the FC layers is that they have a greater number of parameters than other layers, raising computing costs and needing identical input images. The primary disadvantage of a fully-connected layer is that it has a large number of parameters that need complicated calculation during training. As a result, we attempt to reduce the number of nodes and connections. The dropout approach can be used to satisfy the deleted nodes and connections [21,82].

3.4. Deep CNN Architectures

Given the widespread usage of CNNs in medical image processing, we discuss common designs and architectural distinctions across the most extensively used models.

3.4.1. U-NET Architectures

U-Net is a CNN developed for biomedical image segmentation at the Computer Science Department of the University of Freiburg [83]. The network is based on convolutional networks. The encoder–decoder design of U-Net is a stable network for medical image segmentation. Moreover, its architecture was modified and extended to work with fewer training images and more precise segmentation. Segmentation of a 512×512 image takes less than a second on a modern GPU [83]. Ronneberger et al. [83] developed a U-Net network which consists of two stages (path contracting and symmetric expanding). Path

contracting is used to capture, while symmetric expanding is used for precise localization. However, the U-Net architecture depends highly on data augmentation techniques. The architecture of U-Net is composed of 3×3 convolutional layers. Maximum of 2×2 follows each pooling layer and the ReLU activation function. Finally, the 1×1 convolutional layer is attached. The encoder–decoder design of U-Net [83] is a sustainably successful network for medical image segmentation. Because volumetric data is more abundant than 2D images in biomedical data analysis, 3D convolution is suggested and considered considerably more successful in fully using the spatial information of 3D images such as CT and MRI. Based on 3D U-Net [83], Fabian [84] made minor changes. He placed first in several medical image segmentation contests, demonstrating that an optimized U-Net can outperform many other innovative designs.

3.4.2. V-Net Architectures

This section refers to the the proposed V-Net structure in [32]. As with the fundamental U-Net design, the network architecture of the V-Net is composed of encoding and decoding components. Thus, it is a derivation of the U-Net architecture, but with a volumetric design that makes it suited for usage in tissues where organs and tumors are difficult to recognize on CT imaging (such as the prostate or kidney) [32]. The convolutional nature of the V-Net architecture enables it to extract features and reduce the resolution by following the correct route. Because conventional pooling approaches sometimes overlook critical features during the segmentation process, V-Net convolutions are employed to overcome this by downsampling the data provided as input and transmitting it to the receiving characteristics derived in the subsequent network layers [40]. Unlike standard neural network architectures for classification, completely convolutional networks [85], along with the U-Net and the V-Net, do not use a flattening process or contain fully connected layers. Rather than that, it employs upsampling techniques that enable the network to produce an image of the same size as the input, which may address segmentation issues. The V-Net structure is composed of two sections: one for downsampling and one for upsampling. All pooling layers created during the downsampling phase are converted to up-convolution layers during the upsampling phase. Additionally, a “contracting path”, as they termed it, is constructed at each layer from the up-sampling to the down-sampling portion in order to concatenate the data. This method enables the network to view high-resolution data again during the up-sampling phase [86]. Each layer on the encoder side of the V-Net architecture is composed of two times the number of feature set computation sections as the preceding layer. The network’s decoder portion is designed to perform two-channel volumetric segmentation. As a result, feature maps are included to assist in obtaining the relevant information. Following each layer in the encoder portion of the network design, a deflection operation is done to increase the size of the entries, followed by the opposite action in the decoder section to decrease the dimensions. The encoder phase’s neural network attributes are passed to the decoder phase. This is schematically depicted by the use of horizontal linkages [83].

3.4.3. Alex-Net Architectures

CNN is available in many different configurations, including Le-Net, Alex-Net, Google-Net, Conv-Net, and Res-Net. Additionally, we chose the Alex-Net design since it is more resilient to difficulties than alternative designs. Alex-Net has an eight-layer architecture, with the first five layers being convolutional and maximum pooling layers, and the latter three being wholly linked to the neural network [87].

3.4.4. Boundary-Aware FCN Architectures

Although U-Net overcomes the challenge of preserving the original information during FCN upsampling via skip-connection, the boundary cannot achieve a reasonable segmentation result due to the blurring of the border and internal pixels. This uncertainty arises because convolution operators might provide comparable values in the voxel feature

map at the tumor boundary, even in the first convolution layer [88]. Shen et al. [89] proposed a boundary segmentation technique based on a Boundary-Aware FCN network. The Boundary-Aware FCN transforms the single network segmentation problem into a multitasking network. The two FCN networks share the downsampling phase but operate independently of one another during the upsampling phase. The two upsampling correspond to distinct segmentation goals; the first segments the tumor territory, while the second segments the tumor border. The two segmentation results are then fused, and the final segmentation result is produced after numerous convolutional layers.

3.4.5. Cascaded Network Architectures

In a cascaded architecture, the outputs of each CNN are concatenated [90]. The output of one CNN becomes a direct input to another in this architecture. The input cascade is utilized to provide extra image channels to the second CNN by concatenating contextual information. This is an improvement over the dual-path approach, which separates multi-scale label prediction from multi-scale label prediction. Concatenation of local pathways is another variant of cascaded design [90]. Instead of concatenating the output of the first CNN with the output of the second CNN's first hidden layer, this architecture concatenates the output of the first CNN with the output of the second CNN's first hidden layer. Segmentation in a hierarchical fashion [91]: Tumor segmentation is related to organ segmentation in that tumor segmentation needs the location of the tumor based on organ segmentation. A cascading network is proposed as a result of this. The first neural network determines the organ in which the lesion is located, a process termed "rough segmentation", and the second neural network determines the precise tumor segmentation. As for the two-stage cascaded tumor segmentation network, the entire network topology is composed of two U-Net networks. To begin, the original image is sent over the first U-Net network, which converts it to a binary image. After multiplying the image by the original, it is delivered to the second U-Net network. The output result is the segmentation result in its entirety [88].

3.5. Deep Learning Uses in Medical Imaging

3.5.1. Classification

Classification of images or exams was one of the first areas where DL significantly contributed to medical image analysis. Typically, in exam classification, one or more photographs (an exam) are used as input, and a single diagnostic variable is used as an output (e.g., disease present or not). Each diagnostic assessment serves as a sample in this environment, and dataset sizes are often less than those used in computer vision (e.g., hundreds/thousands of samples vs. millions of samples). Transfer learning's appeal for such applications is unsurprising. Transfer learning is the process of utilizing pre-trained networks (usually on natural imagery) in order to circumvent the (perceived) demand for big data sets for deep network training. Two ways of transfer learning were identified: (1) employing a pre-trained network as a feature extractor, and (2) fine-tuning a pre-trained network using medical data. The former technique also has the advantage of avoiding the need to train a deep network, which enables the extracted features to be simply integrated into existing image analysis pipelines. Both tactics are frequently used and popular [27].

3.5.2. Registration

Registration (or spatial alignment) of medical images is a common image analysis task that entails computing the coordinate transform between two medical images. This is frequently done iteratively, with a specified (non-)parametric transformation assumed and a predefined metric (e.g., the L2-norm) optimized. While segmentation and lesion detection are more important topics in deep learning, researchers have demonstrated that deep networks can assist in reaching the highest possible registration performance. In general, two techniques are prevalent in the current literature: (1) employing deep learning networks to quantify the similarity between two images in order to drive an

iterative optimization strategy, and (2) directly predicting transformation parameters using deep regression networks [27].

3.5.3. Segmentation

Segmentation is a typical job in both natural and medical image analysis, and CNNs can easily identify each pixel in the image independently by providing it with patches extracted from surrounding the pixel. A disadvantage of this naive sliding window technique is that input patches from adjacent pixels overlap significantly, resulting in several computations of the same convolution. Because both convolution and dot products are linear operations, and inner products may be represented as convolutions and vice versa. The CNN can process bigger input images than those used for training and generate a probability map rather than a single pixel output by rewriting the fully linked layers as convolutions. The resultant “fully convolutional network” (fCNN) may be applied efficiently to an entire input image or volume [27].

On the other hand, pooling layers may result in output with a far lower resolution than the input. Shift-and-stitch [27] is one of the proposed ways to prevent this loss of resolution. On shifted versions of the input image, the fCNN is applied. A full-resolution version of the final output is obtained by sewing the result together, minus the pixels lost due to the “valid” convolutions. Ronneberger et al. [83] undertook the fCNN concept a step further and suggested the U-Net architecture, which consists of a regular fCNN followed by an upsampling section where up-convolutions are utilized to expand the image size. He invented the terms “contractive” and “expansive” routes. Milletari et al. [32], instead of the standard cross-entropy, presented a modification to the U-Net architecture that integrates ResNet-like residual blocks and a Dice loss layer that directly reduces this extensively used segmentation error metric.

3.5.4. Segmentation Evaluation

Image segmentation has a wide range of applications in most areas of digital image processing. As a result, segmentation evaluation will be crucial in a wide range of disciplines [92]. Manual segmentation defines the boundaries of the Region of Interest (ROI). The bulk of studies in the literature employ manual segmentation to estimate renal volume [93,94]. To check if the segmentation method is valid, first establish the true boundaries of the region of interest. Unfortunately, the major difficulty with medical image segmentation approaches are the lack ground truth. Manually segmenting the Region of Interest (ROI) from the image and comparing those ROIs to ROIs generated by the segmentation method regarding boundary differences or overlap, is a typical solution to this problem [1]. The accuracy of the proposed kidney segmentation method is assessed by comparing the ground truth marked by professional radiologists with the output of the recommended algorithm. Measures for evaluation may be divided into two categories: quantitative and qualitative metrics. A quantitative review includes obtaining mathematical values, whereas a qualitative evaluation requires visually comparing the ground reality with the silhouette generated [1]. The ground truth and the results are compared using error measures. The following are some of the metrics that have been utilized in the literature: Dice Similarity Coefficient (DSC), Specificity (SP), Sensitivity (SN), Accuracy (AC), Jaccard index (JI), Hausdorff Distance (HD), Area Overlap (AO), Area Overlap Error (AOE), Surface Distance (SD), Volume Overlap (VO), Relative Volume Difference (RVD), and Volume Overlap Error (VOE) [95–97]. Some of the various metrics used in the literature are presented in Table 1.

Table 1. Evaluation metric for segmentation.

Metric	Equation	Description
True Positive rate (TP)	$TPR = Sensitivity = Recall = \frac{TP}{(TP+NF)}$	True positive rate, The proportion of true positives or successes that is accurately detected, is calculated as true positive rate, also known as sensitivity [98].
True Negative rate (TN)	$TPR = Sensitivity = Recall = \frac{TP}{(TP+NF)}$	The true negative rate, also known as specificity, is the true negative rate. Is there a chance that a non-diseased will the person be classified as negative by the test? It demonstrates the test's sensitivity in identifying the absence of illness [99].
False-positive rate (FP)	$FP = \frac{FP}{(FP+TN)}$	The false-positive rate refers to the percentage of mistakenly classified as positive or successful but as negative [98].
Dice Similarity Coefficient (DSC)	$DS = 2 \times \frac{ S_m \cap S_a }{ S_m + S_a }$	The binary mask produced by the domain experts' manual segmentation corresponds to the binary mask produced by the suggested approach. DC must be close to unity to guarantee that the manually drawn region corresponds to the segmented result correctly [99].
Jaccard Index JI	$JI = \frac{ S_m \cap S_a }{ S_m + S_a }$	The Jaccard Index (JI) was used to compare the statistical similarity of regions segmented using a computational approach to hand delineations [100,101].
Accuracy	$Accuracy = \frac{(TP+TN)}{(TP+FP+NF+TN)}$	The correct predictions produced by the prediction model across all suitable forecasts completed are referred to as the model's accuracy [102].
Precision	$Precision = \frac{TP}{(TP+FP)}$	The number of correct positive scores divided by the number of positive scores anticipated by the classification algorithm is the positive predictive value, or precision [103].
Sørensen–Dice	$DSC = \frac{2TP}{2TP+FP+FN}$	This coefficient: Indicates the extent to which segmented and reference volumes overlap in mm3. (1 for ideal segmentation, 0 for the worst-case scenario) When applied to Boolean data, the terms true positive (TP), false positive (FP), and false negative (FN) are used (FN), As in this case [47].

3.6. Datasets

There has been substantial scientific interest in automated kidney and renal tumor segmentation during the last few years. As research output increased, objective assessment of various algorithms got more difficult due to researchers' usage of private datasets with various features. Figure 7 and Table 2 summarize the most frequently used datasets for segmenting kidney tumors.

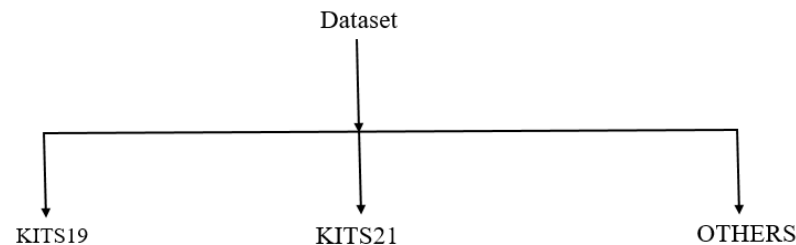


Figure 7. Type of dataset in papers surveyed.

Table 2. Summary of commonly used public datasets for kidney and kidney tumor segmentation.

References	Total	Training Data	Validation	Testing Data
KITS19 [39,40,44,59,72]	300	210	-	90
KITS19 [12]	210	134	34	42
KITS19 [61]	300	240	-	60
KITS19 [62]	300	190	20	90
KITS19 [70]	300	240	30	30
KITS19 [41,58]	300	168	42	90
KITS21 [55,63,65,69]	300	240	-	60
OTHER [50]	113	70	23	20
OTHER [54]	140	90	-	50

The KiTS19 and KiTs21 Challenges and additional datasets provide a sizable dataset that enables a variety of segmentation tasks. The challenge intends to further research on general-purpose segmentation algorithms capable of performing a variety of tasks without the need for human assistance for kidney segmentation and tumors.

3.7. Techniques for Kidney Tumor Segmentation

3.7.1. Pre-Processing

Typically, data is preprocessed before to being transmitted to the deep learning network. Preprocessing is grouped into four categories: approach based on image attributes, data augmentations, noise removal, and edge improvement. They have essentially agreed that if the image is placed directly into the deep neural network without preprocessing, the effect is considerably reduced, and in some cases, proper preprocessing is critical to the model's performance [88,104]. Among the different bias field correction strategies are the non-parametric non-uniform normalization (N3) approach [105]. It has emerged as the preferred methodology owing to its ease of use and availability as an open-source project [106]. This approach was further enhanced in [106], and it is now often referred to as N4ITK. These approaches are intended to be used with a single image. As a result, the intensity normalization provided by Nyul et al. [107] may be used to achieve a consistent intensity distribution across patients and acquisitions. Another often used preprocessing approach is normalizing the image dataset such that it has a mean of zero and a standard deviation of one. This approach aids in the de-biasing of characteristics. Cropping an image may also be used to eliminate as many background pixels as feasible.

3.7.2. Post-Processing

The output of the deep neural network may be used directly as the result of tumor segmentation. However, the output of a deep neural network is not always immediately applicable to the demands and is not interpretable. As a result, some researchers perform post-processing techniques on the output of DL to obtain more accurate findings. Post-processing is used to fine-tune the segmentation findings. It aids in reducing the number of misclassifications or false positives in segmentation results when using algorithms such as conditional random fields (CRF) [108] and Markov random fields [32]. CRF and MRF-based techniques effectively eliminate false positives by combining model predictions with low-level image information, such as local interactions of pixels and edges, when performing finer adjustments. These approaches, however, are computationally intensive [109]. Related component analysis entails locating and identifying connected components and removing unnecessary blobs using a simple thresholding approach. Another strategy for reducing false positives around the segmentation image's boundaries is to perform successive morphological operations such as erosion and dilation.

3.7.3. Data Augmentation

The purpose of data augmentation, which is to increase the number of training sets, is to expand the size of the data set by graphical modification, hence making the model more robust and less prone to overt error. Flip, Rotation, Shift, Shear, Zoom, Brightness, and Elastic distortion are all common strategies for increasing the number of datasets [110]. To emphasize that when dealing with a limited data collection, the advantage of data augmentation is equivalent to the benefit of model update [88]. However, it is easy to implement. The problem of lack of access to huge data in kidney and kidney tumors can be solved using Data Augmentation.

4. Overview of Kidney Tumor Semantic Segmentation

This section provides a quick overview of kidney tumor segmentation. Renal segmentation is a complex operation; the difficulty of segmentation varies depending on the imaging modality. Various imaging techniques are available.

4.1. Renal Imaging

The size of the kidney was established in the past by X-rays or using renal length for urography. The results obtained using these methods showed several problems. Ultrasound sonography (US), computed tomography (CT), and magnetic resonance imaging (MRI) are examples of photography methods that may be used to examine the size and function of the kidneys. It is the first method (US) for measuring two-dimensional kidney volumes. To obtain 3D data, they employed CT and MRI techniques. Each of the current imaging modalities has its imaging capabilities and may be used in various ways depending on the treatment objective [111]. US can identify cysts, stones, and tumors, among other benefits; it gives good anatomical information without exposing the patient to radiation and enabling a low-cost, real-time inspection. However, the images on US are of poor quality [111]. This is a flaw that makes the segmentation procedure difficult. On the other hand, computed tomography (CT) is a technique that allows for higher-quality imaging and identifies small lesions and cysts. However, there is a drawback to ionizing radiation exposure. The last approach (MRI) is insufficient. Overall, the benefit of MRI is that it provides excellent spatial resolution while posing minimal risk to the patient. Its downside is that it is more expensive [111].

4.2. Image Segmentation

What exactly is an image? An image is a mathematical representation of what we can see. An image can be defined as a 2D function like a CT image denoted by $f(x, y)$, where the value or amplitude off at spatial coordinates (x, y) gives the intensity (brightness) of

the image at that point. A pixel is a visual element that is represented by each end in an image. The function f can also be viewed here:

1. The size of the image array $N \times X, A$ where
2. S and M represent the number of rows and columns.

Thus,

$$A = f(x, y) = \begin{pmatrix} f_{(0,0)} & f_{(0,1)} & \cdots & f_{(0,M-1)} \\ f_{(1,0)} & \vdots & \cdots & f_{(1,M-1)} \\ \vdots & \vdots & \cdots & \vdots \\ f_{(N-1,1)} & f_{(N-1,1)} & \cdots & f_{(N-1,M-1)} \end{pmatrix} \quad (1)$$

Image segmentation, defining objects in images, is the most important operation performed on acquired images. Segmentation, ironically, is required for segmentation, as object knowledge helps with segmentation. In image processing and computer vision, segmentation is still a difficult task [112]. The assessment of segmentation algorithms is a closely connected and strongly interrelated topics. The absence of a uniform framework for evaluating segmentation algorithms is one of the challenges in creating them [112]. The initial stage in creating a wholly automated perception system is image segmentation. It has been regarded as a primary challenge in computer vision. Generally speaking, the mission of segmenting an image into parts might be to identify objects or districts of interest [113]. The following section explains the types of image segmentation:

4.3. Types of Segmentation

The degree of human participation in image segmentation algorithms may be classified into four types: manual segmentation, semi-automatic segmentation, fully automatic segmentation, and semantic segmentation [114]. A summary of the advantages and disadvantages of segmentation techniques is given in Table 3.

Table 3. Summary of the benefits and drawbacks of various segmentation techniques.

Type of Segmentation	Reproducibility	Time	Interactivity	Complexity of Implementation
Manual Segmentation	Good	Too long	Bad	Easy
Semi-Automatic Segmentation	Good	Long	Not bad	Easy
Fully Automatic Segmentation	Good	Short	Good	Hard
Semantic Segmentation	Good	Short	Good	Hard

4.3.1. Manual Segmentation

Manual segmentation is done drawing the member or sketching the member's borders, manually executed by a professional (physician, trained, technician, etc.) [115]. The person operating the machine uses specialized tools to draw or paint around tumor regions. Manual segmentation is a time-consuming and tiresome process. The result of segmentation depends on the training and experience of the person. It also includes what is called "inter-rater" variability (the image is segmented differently by various specialists) and "intra-rater" variability (at various times, the same person segments the image differently). It serves as the basis for various segmentation algorithms (semi-automatic, automatic, and semantic). Even though manual segmentation is dependent on the rater, it is still commonly utilized in clinical studies [116–118].

4.3.2. Semi-Automatic Segmentation/Interactive Segmentation

To address manual segmentation problems, semi-automated or automatic segmentation approaches have been developed. Semi-automated segmentation approaches necessitate the operator's one-time manual startup and manual modification of the computer segmentation result [1]. Semi-Automated Segmentation (SAS) takes less time than manual segmentation, but its outcomes are still reliant on the operator [40,114,115].

4.3.3. Fully Automatic Segmentation

To have as little human interaction as possible, automatic segmentation methods involve the segmentation of kidneys from CT images without human intervention, making them free from human errors and biases [1]. Utilizing anatomical knowledge such as volume, form, and position in automated segmentation algorithms is critical for creating a robust algorithm [40,114]. Radiologists like this sort of approach since it is error-free and operator-independent. Several types of segmentation approach in the literature may be classified as thresholding-based methods, region-based methods, model-based methods, atlas-based methods, and hybrid methods depending on the strategy used to segment the region of interest. Depending on the user's needs, these segmentation approaches can be automated or semi-automatic [1].

4.3.4. Semantic Segmentation

Semantic segmentation, also known as pixel-level classification, aims to group portions of images corresponding to the same object class. This sort of algorithm may be used to recognize road signs, cancers, and medical tools in surgeries, among other things [40]. Semantic segmentation is a particular job that attempts to split an image into semantically meaningful pieces, making it a step farther than image segmentation. (It is important to remember that semantics is a discipline of linguistics concerned with meaning) [119]. It is a high-level task that paves the way for complete scene comprehension in the broad image. The importance of scene understanding as a crucial computer vision problem is underscored by the fact that an increasing number of applications rely on inferring knowledge from images. Classification of an image refers to assigning it to one of the same categories. Detection is the process of locating and recognizing objects. Because it classifies each pixel into its category, image segmentation may be considered pixel-level prediction. In addition, there is a job called instance segmentation that combines detection and segmentation [120].

4.3.5. Semantic Segmentation Metrics

Semantic segmentation metrics play an essential role in evaluating performance techniques. Different semantic segmentation assessment criteria may produce disparate results because it is unclear how to define successful performance segmentation. Pixel accuracy, mean intersection over union, and representing per-class accuracy are three of the most frequently used measures [121].

For all of them, let $n_{i,j}$ be the number of class, pixels predicted to belong to the classes, j In addition, let $k_j = \sum_i n_{i,j}$ be the total pixel number belonging to class i . If we assume to have a T total number of classes, then:

- Pixel accuracy can be computed as:

$$acc = \frac{\sum_j n_{i,j}}{\sum_i n_{i,j}}$$

- Mean intersection over union can be computed as:

$$miou = \frac{1}{T} \sum_i \frac{n_{i,i}}{(k_i + \sum_j n_{j,i} - n_{i,i})}$$

- Mean per class accuracy can be computed as:

$$miou = \frac{1}{T} \frac{\sum_i n_{i,j}}{\sum_i k_i}$$

5. Discussion

Deep learning algorithms for medical image processing have garnered considerable attention in recent years. This is reflected in the year-over-year growth in the number of published works [122]. Recent years have seen the use of LD approaches to a wide variety of problems [123], most notably in computer vision [124], natural language processing [125], and computational audio analysis [126]. DL algorithms have outperformed past state-of-the-art performance in a number of these applications. The domain-independent concept of effectively accomplishing high-level tasks via hierarchical layers of learned abstraction [123]. For instance, DL algorithms can be trained on a sufficiently big dataset to segment TC images to segment kidney tumors. As a result of their effectiveness in tackling numerous problems in computer vision, speech recognition, and natural language processing, CNN-based models have been widely utilized in medical image analysis. Table 4 summarizes the deep learning approaches discussed in this paper. Numerous techniques have significant differences in terms of architectural design, with current works following the U-Net [83]. Additionally, numerous approaches have been developed to solve the inherent difficulties associated with semantic renal TC analysis. Table 5 gives an overview of Deep Learning methods for kidney tumor segmentation on other architecture.

Table 4. Overview of Deep Learning methods for kidney tumor segmentation: PA = Pixel accuracy, SS = Specificity-Sensitivity, KD = Kidneys Dice, TD = Tumor Dice, CD = Composite Dice, DSC = Dice = Dice similarity coefficient DSC, CD = Centroid distance, HD = Hausdorff distance, AC = Accuracy, SGD = Stochastic gradient descent, SD = Surface Dice, BN = Batch Normalization, IN = Instance Normalization, SGD = Stochastic Gradient Descent, IOU = Intersection over Union.

Reference	Input	Regulization	Activation	Loss	Optimizer	
U-Net Architecture						
[127]	3D	BN,Depthwise, Weight Pruning		Dice	Decathlon	
[128]	3D		RELU	Mean IoU, AC	Adam	
[49]	3D,2D		RELU		Dice	SGD
[129]	3D				Dice	Adam
[60]	3D			Leaky ReLU	Dice	Adam
[53]	3D			Dice, SD	Adam	
[12]	3D	IN		Dice	Adam	
[130]	3D,2D	BN	ReLU	Dice	Adam	
[131]		BN	RELU	IOU		
[56]	3D		ReLU	Dice	Adam	
[55]	3D	BN	ReLU	Dice	Adam	
[69]	3D		ReLU	Dice	Adam	
[68]	3D	Batch norm	ReLU	Dice	Adam	
[67]	3D	BN	ReLU	Dice	Adam	
Cascaded Architecture						
[63]	2.5D	BN	ReLU,conv	Dice	Adam	
[41]	3D	BN	SE-Net	Dice	Adam	
[51]		Dropout	RELU	Dice, CD, HD		
[59]	3D	BN	ReLU	Dice	Adam	
[65]	2D	BN	RELU, LeakyRelu	Dice		
3D U-Net Architecture						
[39]	3D		RELU	Dice	SDG	
[66]	3D		RELU	Dice		
[132]	3D	BN	ReLU	Dice		
[57]	3D	BN	ReLU	Dice	Adam	

Table 4. *Cont.*

Reference	Input	Regulization	Activation	Loss	Optimizer
Boundary-Aware Architecture					
[44]	3D	BN	RELU	Dice	Adam
[5]	3D		RELU	KD, TD, CD	Adam
V- Net Architecture					
[40]	3D			Dice	Adam
[19]				Dice	
Ensemble Architecture					
[47]	2D		RELU	Dice	Adam
[64]				Dice	

Table 5. Overview of Deep Learning methods for kidney tumor segmentation on other architecture.

Reference	Architecture	Input	Regulization	Activation	Loss	Optimizer
[50]	MB-FSGAN	3D	BN	RELU	PA, Dice, SS	RMSProp, Adam
[58]	U-Net, AlexNet	2D	BN	RELU	Dsc, Jaccard index, AC, SS	Adam
[133]	Modified CNN	2D	Weight Decay		Dice	
[61]	EG-CNN	3D		RELU	Dice	Adam
[54]	FCN	3D	L2		Dice	SDG
[46]	RAU-Net	3D			Dice	SDG
[62]	multi-stage U-Net	2.5D	BN	pre-activation	Dice	Adam
[52]	CTumor GAN	3D	BN, Dropout	RELU	Dice, Jaccard index, SS	Adam
[73]	nnU-Net	3D	IN		Dice, Jaccard, Ac, Precision, Recall, Hausdorff	Adam
[48]	FPN (CNN)	2D			Dice	
[45]	CNN	2D,3D			Dice	
[71]	3D SEAU-Net	3D	BN		Dice	
[134]	DeepLab v3+	3D	BN	RELU	Dice	Adam

5.1. Kidney Tumor

While the techniques listed below are quite effective in segmenting kidney tumors, none of them quantifies kidney tumors. CT scans are frequently used to diagnose kidney malignancies, and deep learning-based segmentation of kidney tumors is also focused on CT images. Yang et al. [54] made one of the earliest attempts to use CNNs for this goal. An enhanced residual fully connected network with a pyramidal pool module is

presented to segment kidneys and kidney cancers in CT angiography images. To begin, an atlas-based approach is used to extract two zones of interest from the entire images. Each area is home to a single kidney. These patches are fed into a 3D CNN that is profoundly linked, and additional post-processing is conducted using conditional random fields. Efremova et al. [45] and Shen et al. [135] applied U-Net and 3D U-Net to the job of kidney tumor segmentation, with all techniques achieving much better results than conventional methods [51]. The next step was to analyze images from two distinct viewpoints and then link them via two connected networks. Crossbar-Net, a new architecture for automatic segmentation of kidney tumors in CT images, was described. The axial slice was created by stacking three rectangular patches. These patch designs included extra spatial information and were used to train two separate networks with cascading outputs. Cascaded network architectures were used by Yang et al. [54], Vu et al. [136], Lv et al. [137], Mu et al. [19], and Wei et al. [64] to discriminate between kidney cancers. The difference between the two methods is that Yang et al. [54] used a Gaussian pyramid to expand the receptive field in the first stage's network structure, while Vu et al. [136] increased the number of layers in the cascade network to three, with the first layer obtaining the results directly, the second layer obtaining the tumor and kidney regions, the third layer obtaining the tumor segmentation results using the input of the second layer cascade, and the fourth layer cascading the final results. The two cascade frameworks in [137] are U-Net and V-Net to achieve distinct cascade characteristics. Recently, Xia et al. [138] suggested a two-stage segmentation strategy for the kidney and the space-occupying lesion region. This approach retrieves images using Spatial CNN (SCNN) and Resnet, and smooths and matches pixels using SIFT-flow and Madras Rubber Factory (MRF). Table 5 shows an overview of Deep Learning methods for kidney tumor segmentation on other architectures. Table 6 shows a summary of results on KiTs 2019, KiTs2021, and another dataset.

Table 6. A summary of results on KiTs 2019, KiTs2021, and another dataset.

Reference	Kidneys Dice	Tumor Dice	Composite Dice
KiTS19			
[61]	0.965	0.835	0.900
[60]	0.967	0.845	0.906
[39]	0.974	0.851	0.912
[129]	0.97	0.32	
[59]	0.974	0.831	0.902
[12]	0.969	0.805	0.887
[62]	0.98	0.73	0.855
[40]	0.977	0.865	0.921
[19]	0.973	0.817	
[70]	0.978	0.868	0.923
[5]	0.974	0.810	0.892
[134]	0.872	0.384	
[41]	0.968	0.743	0.856
[47]	0.949	0.601	
[44]	0.970	0.834	0.902
[46]	0.960	0.770	
[66]	0.930	0.570	
[64]	0.968	0.750	
[45]	0.964	0.674	
[71]	0.924	0.743	
[72]	0.852		

Table 6. Cont.

Reference	Kidneys Dice	Tumor Dice	Composite Dice
KiTS21			
[63]	0.943	0.778	0.871
[49]	0.975	0.881	
[53]	0.923	0.553	
[65]	0.934	0.643	
[67]	0.96	0.81	
[68]			0.654
[69]	0.916	0.541	825
[55]	0.937	0.750	
[56]	0.90	0.39	
Other Dataset			
[50]			0.859
[54]	0.923	0.826	0.875
[51]			0.925

While the method outlined above is quite effective in segmenting kidneys, the consequence of a kidney tumor is clearly different for papers using KiTs19. As seen in Figure 8, the V-Net and U-Net are the optimal configurations. The comparison using KiTs21 and the other dataset are shown in Figure 9. Table 7 shows a summary of results using other metrics.

Table 7. A summary of results using other metrics.

Reference	Sensitivity	Specificity	Jaccard	Accuracy	Hausdorff
[12]			0.716	0.99	33.469
[70]	0.913	0.914			5.10
[50]	0.862	0.894		0.957	
[72]	0.842	0.998	0.756	0.997	18.39

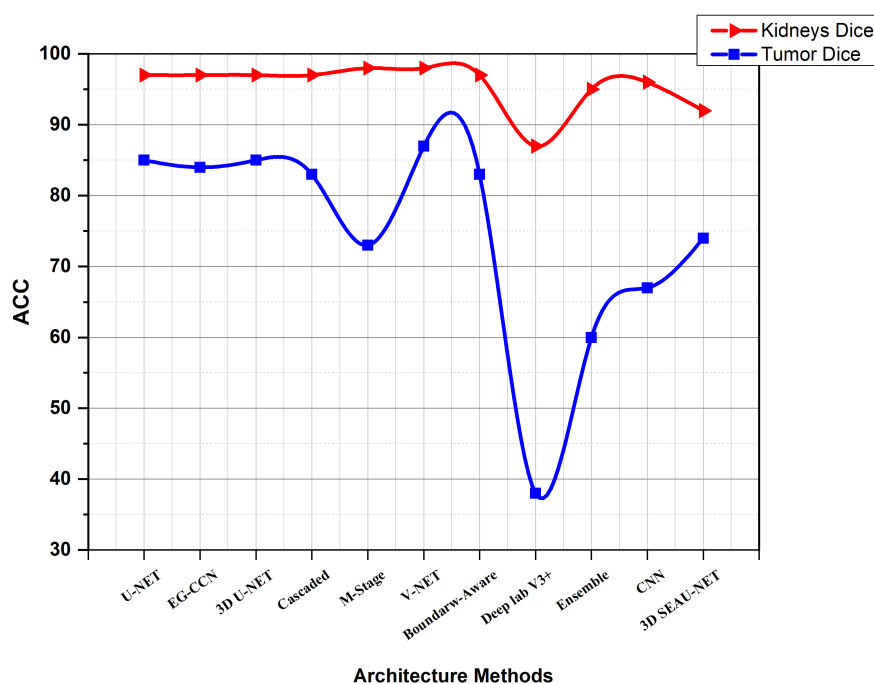


Figure 8. Diagram showing a comparison between different Architecture Methods using KiTs19.

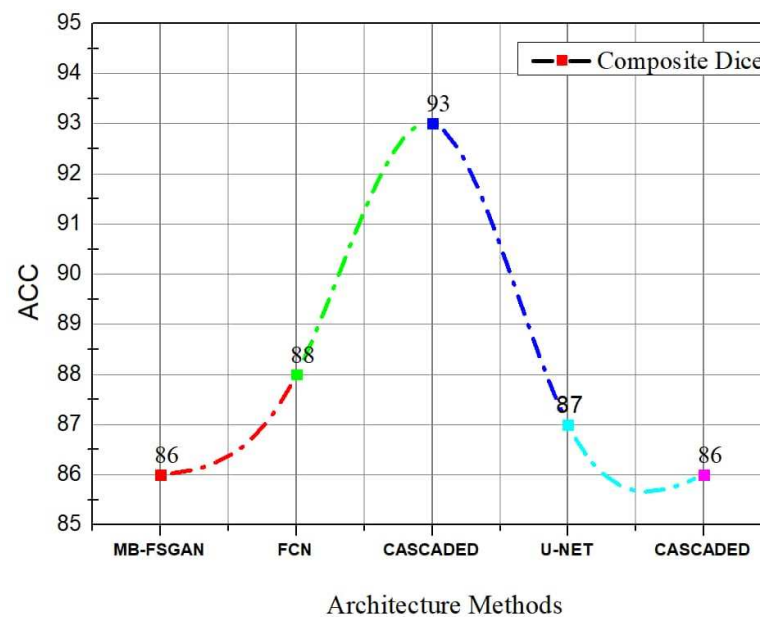


Figure 9. Diagram showing a comparison between different Architecture Methods using KiTs21 and another dataset.

5.2. Deep Learning

The DL approach is trained to do segmentation on a pixel-by-pixel or region-by-region basis. To begin, features are retrieved to distinguish the lesion from the background, and different lesion characteristics can be integrated via feature extraction. However, due to the lengthy training period and potential for over-training, test images should originate from the same platform as the training images. DL techniques require a huge quantity of training data to perform effectively on previously unknown materials. However, this presents several difficulties in the medical area. First, annotating even a single CT volume requires a substantial amount of time for a well-trained radiologist. Additionally, the work is prone to intra- and inter-rater variability. As a result, all annotations are authorized by an excessive number of expert neuroradiologists [139].

A significant amount of effort is spent obtaining permission, and fortunately, deep convolutional neural networks have demonstrated exceptional performance in a variety of computer vision applications. These networks, however, rely largely on massive amounts of data to avoid overfitting. Overfitting is a term that refers to the process by which a network learns a function with a very large variance in order to predict the training data properly. Regrettably, many application fields, such as medical image analysis, lack access to huge data [140]. Because the unavailability of large-scale datasets limits the potential of deep learning models, researchers have turned to data augmentation as a quick fix for the data issues described above. Other recent studies have examined unsupervised learning [141]. Numerous augmentations have been proposed, most of which can be categorized as data warping or oversampling techniques. Data augmentation has a very bright future. The potential for using search algorithms that combine data warping and oversampling techniques is immense. Deep neural networks' layered architecture provides numerous chances for data enhancement [81].

Regarding hardware requirements while training a model with huge datasets, DL techniques demand massive computations processes. Due to the fact that the advantage of a GPU over a CPU rises with the scale of the computations, the GPU is mostly used to successfully optimize processes. Thus, GPU hardware is required to perform successfully with deep learning training. As a result, DL is more reliant on high-performance machines equipped with GPUs than on traditional machine learning approaches [142]. Additionally, the use of 3D deep learning models significantly increases the computational and memory

requirements. DL uses software libraries to define and train deep neural networks in a parallel or distributed manner while leveraging multi-core or multi-GPU environments. At the moment, researchers are constrained by the amount of GPU memory available to them (typically 12 gigabytes). As a result, batch sizes and model complexity are constrained by the amount of memory available [21].

The performance of kidney tumor segmentation algorithms have continued to improve in recent years as more training data has been available and more sophisticated CNN architectures and training techniques have been used. Their robustness, however, continues to fall behind expert performance [21]. Interestingly, single U-Net [83] based models [143] continue to perform well, corroborating the assertion that “a well-trained U-net is difficult to surpass” [144]. As a result, concerns such as properly analyzing 3D slice data and compressing the model as the number of network parameters rises, must be addressed in the future. Additionally, because of the scarcity of medical imaging data, researchers must consider how to avoid over tagging. Although the features of individual tumors vary, there are several similarities between them. However, no successful study presently aims to unify segmentation algorithms and transfer learning across tumor types [88]. The examined literature demonstrates that careful initialization of hyper-parameters, pre-processing techniques, sophisticated training schemes, and addressing the class imbalance problem contribute significantly to the accuracy and resilience of segmentation algorithms.

6. Conclusions and Future Work

This paper discusses ways of segmenting kidneys and kidney tumors using deep learning and building blocks, as well as state-of-the-art approaches and implementation tools. The existing techniques serve two purposes: segmenting tumors correctly and compensating for the lack of training data. Based on adequate training data, DL is capable of adequately segmenting kidney tumors. With proper pre-processing, weight initialization, sophisticated training schemes, segmentation with unambiguous borders, and obtaining additional information for pixel classification, ensemble approaches and U-Net-based models have significant potential for improving the state-of-the-art. The absence of a large-scale medical training dataset is a primary reason for the poor performance of many segmentation algorithms. As a starting point for future development, overall, kidney and renal tumor segmentation challenges have been met with great success. It received a large number of submissions and continues to be a significant and hard benchmark for 3D segmentation. However, extending the use of these systems outside the sampled population for the test set would be desirable since it was obtained from individuals who shared the same geographic region and healthcare system and a multi-institutional cohort with a prospectively generated test set. Additional imaging modalities, such as magnetic resonance imaging (MRI) or contrast-enhanced ultrasound (CEUS), may be employed to increase the diagnostic algorithm’s accuracy when CT alone is used. Future research on DL designs, particularly in the realm of medical imaging, should avoid complicated architectures. Future research should focus on reducing the training time for deep learning models. To produce more successful models that can be utilized in various disciplines, it is required to simplify the systems to which the models may be applied and minimize their complexity.

Author Contributions: Contributions: Conceptualization, A.A. and S.V.; methodology, A.A.; formal analysis, A.A. and S.V.; investigation, A.A.; resources, S.V.; writing original draft preparation, A.A.; writing review and editing, S.V.; supervision, S.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data available in publicly accessible repositories.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kaur, R.; Juneja, M. A Survey of Kidney Segmentation Techniques in CT Images. *Curr. Med. Imaging Rev.* **2017**, *14*, 238–250. [\[CrossRef\]](#)
2. Chen, X.; Summers, R. M.; Yao, J. Kidney Tumor Growth Prediction by Coupling Reaction-Diffusion and Biomechanical Model. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 169–173. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Du, Z.; Chen, W.; Xia, Q.; Shi, O.; Chen, Q. Trends and Projections of Kidney Cancer Incidence at the Global and National Levels, 1990–2030: A Bayesian Age-Period-Cohort Modeling Study. *Biomark. Res.* **2020**, *8*, 1–10. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Smith, C.F.; Harvey, L.; Mills, K.; Harrison, H.; Rossi, S. H.; Griffin, S. J.; Stewart, G. D.; Usher, J. A. Reasons for Intending to Accept or Decline Kidney Cancer Screening: Thematic Analysis of Free Text from an Online Survey. *BMJ Open* **2021**, 1–12. [\[CrossRef\]](#)
5. Myronenko, A.; Hatamizadeh, A. 3D Kidneys and Kidney Tumor Semantic Segmentation Using Boundary-Aware Networks. *arXiv* **2019**, arXiv:1909.06684.
6. Heller, N.; Sathianathan, N.; Kalapara, A.; Walczak, E.; Moore, K.; Kaluzniak, H.; Rosenberg, J.; Blake, P.; Rengel, Z.; Oestreich, M.; et al. The KiTS19 Challenge Data: 300 Kidney Tumor Cases with Clinical Context, CT Semantic Segmentations, and Surgical Outcomes. *arXiv* **2019**, arXiv:1904.00445.
7. Grief, B.; Eisen, T. Medical Treatment of Renal Cancer: New Horizons. *Br. J. Cancer* **2016**, *115*, 505–516. [\[CrossRef\]](#)
8. Hua, X.; Shi, H.; Zhang, L.; Xiao, H.; Liang, C. Systematic Analyses of The Role of Prognostic And Immunological of EIF3A, A Reader Protein, in Clear Cell Renal Cell Carcinoma. *Cancer Cell Int.* **2021**, *21*, 118.
9. Millet, I.; Doyon, F.C.; Hoa, D.; Thuret, R.; Merigeaud, S.; Serre, I.; Taourel, P. Characterization of Small Solid Renal Lesions: Can Benign and Malignant Tumors Be Differentiated with CT? *Am. J. Roentgenol.* **2011**, *197*, 887–896. [\[CrossRef\]](#)
10. Chawla, S.N.; Crispen, P.L.; Hanlon, A.L.; Greenberg, R.E.; Chen, D.Y.T.; Uzzo, R.G. The Natural History of Observed Enhancing Renal Masses: Meta-Analysis and Review of the World Literature. *J. Urol.* **2006**, *175*, 425–431. [\[CrossRef\]](#)
11. Xie, Y.; Ma, X.; Li, H.; Gao, Y.; Gu, L.; Chen, L.; Zhang, X. Prognostic Value of Clinical and Pathological Features in Chinese Patients with Chromophobe Renal Cell Carcinoma: A 10-Year Single-Center Study. *J. Cancer* **2017**, *8*, 3474. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Zhao, W.; Jiang, D.; Peña Queralta, J.; Westerlund, T. MSS U-Net: 3D Segmentation of Kidneys and Tumors from CT Images with a Multi-Scale Supervised U-Net. *Inform. Med. Unlocked* **2020**, *19*, 100357. [\[CrossRef\]](#)
13. Pan, X.; Quan, J.; Li, Z.; Zhao, L.; Zhou, L.; Jinling, X.; Weijie, X.; Guan, X.; Li, H.; Yang, S.; Gui, Y.; Lai, Y. MiR-566 Functions as an Oncogene and a Potential Biomarker for Prognosis in Renal Cell Carcinoma. *Biomed. Pharmacother.* **2018**, *102*, 718–727. [\[CrossRef\]](#)
14. Hollingsworth, J.M.; Miller, D.C.; Daignault, S.; Hollenbeck, B.K. Rising Incidence of Small Renal Masses: A Need to Reassess Treatment Effect. *J. Natl. Cancer Inst.* **2006**, *98*, 1331–1334. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Cairns, P. Renal Cell Carcinoma. *Cancer Biomark.* **2011**, *9*, 461–473. [\[CrossRef\]](#) [\[PubMed\]](#)
16. McLaughlin, J. K. Epidemiologic Characteristics and Risk Factors for Renal Cell Cancer. *Clin. Epidemiol.* **2009**, *1*, 33. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Sun, M.; Abdollah, F.; Bianchi, M.; Trinh, Q. D.; Jeldres, C.; Thuret, R.; Tian, Z.; Shariat, S. F.; Montorsi, F.; Perrotte, P.; Karakiewicz, P. I. Treatment Management of Small Renal Masses in the 21st Century: A Paradigm Shift. *Ann. Surg. Oncol.* **2012**, *19*, 2380–2387. [\[CrossRef\]](#)
18. Journal, P. Noise Issues Prevailing in Various Types of Medical Images. *Biomed. Pharmacol. J.* **2018**, *11*, 1227–1237.
19. Mu, G.; Lin, Z.; Han, M.; Yao, G.; Gao, Y. Segmentation of Kidney Tumor by Multi-Resolution VB-Nets. *Univ. Minn. Libr.* **2019**, 1–5. [\[CrossRef\]](#)
20. Jung, I.K.; Jeong, Y.C.; Kyung, C.M.; Hak, J.L.; Seung, H.K. Segmental Enhancement Inversion at Biphasic Multidetector CT: Characteristic Finding of Small Renal Oncocytoma. *Radiology* **2009**, *252*, 441–448. [\[CrossRef\]](#)
21. Magadza, T.; Viriri, S. Deep Learning for Brain Tumor Segmentation: A Survey of State-of-the-Art. *J. Imaging* **2021**, *7*, 19. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Dass, R.; Devi, S. Image Segmentation Techniques. *Comput. Vis. Graph. Image Process.* **2012**, *7109*, 66–70.
23. Lateef, F.; Ruichek, Y. Survey on Semantic Segmentation Using Deep Learning Techniques. *Neurocomputing* **2019**, *338*, 321–348. [\[CrossRef\]](#)
24. Xu, W.; Li, B.; Liu, S.; Qiu, W. Real-Time Object Detection and Semantic Segmentation for Autonomous Driving. *MIPPR* **2018**, *2020*, 44. [\[CrossRef\]](#)
25. Tseng, Y.H.; Jan S.S. Combination of Computer Vision Detection and Segmentation for Autonomous Driving. *IEEE/ION Position Locat. Navig. Symp. Plans* **2018**, 1047–1052. [\[CrossRef\]](#)
26. Jiang, F.; Grigorev, A.; Rho, S.; Tian, Z.; Fu, Y. S.; Jifara, W.; Adil, K.; Liu, S. Medical Image Semantic Segmentation Based on Deep Learning. *Neural Comput. Appl.* **2018**, *29*, 1257–1265. [\[CrossRef\]](#)
27. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A Survey on Deep Learning in Medical Image Analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [\[CrossRef\]](#)
28. Hesamian, M.H.; Jia, W.; He, X.; Kennedy, P. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *J. Digit. Imaging* **2019**, *32*, 582–596. [\[CrossRef\]](#)

29. Liu, H.; Lang, B. Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey. *Appl. Sci.* **2019**, *9*, 4396. [CrossRef]
30. Hatamizadeh, A.; Hosseini, H.; Liu, Z.; Schwartz, S. D.; Terzopoulos, D. Deep Dilated Convolutional Nets for the Automatic Segmentation of Retinal Vessels. *arXiv* **2019**, arXiv:1905.12120.
31. Hatamizadeh, A.; Hoogi, A.; Sengupta, D.; Lu, W.; Wilcox, B.; Rubin, D.; Terzopoulos, D. Deep Active Lesion Segmentation. *Lect. Notes Comput. Sci.* **2019**, *11861*, 98–105. [CrossRef]
32. Milletari, F.; Navab, N.; Ahmadi, S.A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. *Int. Conf. 3D Vision* **2016**, *2016*, 565–571. [CrossRef]
33. Lo, S.-C.B.; Lou, S.-L.A.; Lin, J.-S.; Freedman, M.T.; Chien, M.V.; Mun, S.K. And Applications for Lung Nodule Detection. *IEEE Trans. Med. Imaging* **1995**, *14*, 711–718. [CrossRef] [PubMed]
34. Christ, P.F.; Ettlinger, F.; Grün, F.; Elshaera, M.E.A.; Lipkova, J.; Schlecht, S.; Ahmaddy, F.; Tatavarty, S.; Bickel, M.; Bilic, P.; et al. Automatic Liver and Tumor Segmentation of CT and MRI Volumes Using Cascaded Fully Convolutional Neural Networks. *arXiv* **2017**, arXiv:1702.05970.
35. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 936–944. [CrossRef]
36. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239. [CrossRef]
37. Doll, P.; Girshick, R.; Ai, F. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
38. Guo, Y.; Liu, Y.; Georgiou, T.; Lew, M.S. A Review of Semantic Segmentation Using Deep Neural Networks. *Int. J. Multimed. Inf. Retr.* **2018**, *7*, 87–93. [CrossRef]
39. Isensee, F.; Maier-Hein, K.H. An Attempt at Beating the 3D U-Net. *arXiv* **2019**, arXiv:1908.02182.
40. Türk, F.; Lüy, M.; Barışçi, N. Kidney and Renal Tumor Segmentation Using a Hybrid V-Net-Based Model. *Mathematics* **2020**, *8*, 1772. [CrossRef]
41. Xie, X.; Li, L.; Lian, S.; Chen, S.; Luo, Z. SERU: A Cascaded SE-ResNeXt U-Net for Kidney and Tumor Segmentation. *Concurr. Comput. Pract. Exp.* **2020**, *32*, 1–11. [CrossRef]
42. Ali, A.M.; Farag, A.A.; El-Baz, A.S. Graph Cuts Framework for Kidney Segmentation with Prior Shape Constraints. *Lect. Notes Comput. Sci.* **2007**, *4791*, 384–392. [CrossRef]
43. Thong, W.; Kadoury, S.; Piché, N.; Pal, C.J. Convolutional Networks for Kidney Segmentation in Contrast-Enhanced CT Scans. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **2018**, *6*, 277–282. [CrossRef]
44. Myronenko, A.; Hatamizadeh, A. *Edge-Aware Network for Kidneys and Kidney Tumor Semantic Segmentation*; University of Minnesota Libraries Publishing: Mankato, MN, USA, 2019.
45. Efremova, D.B.; Konovalov, D.A.; Siriapisith, T.; Kusakunniran, W.; Haddawy, P. Automatic Segmentation of Kidney and Liver Tumors in CT Images. *arXiv* **2019**, arXiv:1908.01279.
46. Guo, J.; Zeng, W.; Yu, S.; Xiao, J. RAU-Net: U-Net Model Based on Residual and Attention for Kidney and Kidney Tumor Segmentation. In Proceedings of the 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 15–17 January 2021; pp. 353–356. [CrossRef]
47. Causey, J.; Stubblefield, J.; Qualls, J.; Fowler, J.; Cai, L.; Walker, K.; Guan, Y.; Huang, X. An Ensemble of U-Net Models for Kidney Tumor Segmentation with CT Images. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* **2021**, *5963*, 1–5. [CrossRef]
48. Nazari, M.; Jiménez-Franco, L.D.; Schroeder, M.; Kluge, A.; Bronzel, M.; Kimiaei, S. Automated and Robust Organ Segmentation for 3D-Based Internal Dose Calculation. *EJNMMI Res.* **2021**, *11*, 1–13. [CrossRef]
49. George, Y. A Coarse-to-Fine 3D U-Net Network for Semantic Segmentation of Kidney CT Scans. Available online: <https://openreview.net/forum?id=dvZiPuZk-Bc> (accessed on 10 October 2021).
50. Ruan, Y.; Li, D.; Marshall, H.; Miao, T.; Cossetto, T.; Chan, I.; Daher, O.; Accorsi, F.; Goela, A.; Li, S. MB-FSGAN: Joint Segmentation and Quantification of Kidney Tumor on CT by the Multi-Branch Feature Sharing Generative Adversarial Network. *Med. Image Anal.* **2020**, *64*, 101721. [CrossRef]
51. Yu, Q.; Shi, Y.; Sun, J.; Gao, Y.; Zhu, J.; Dai, Y. Crossbar-Net: A Novel Convolutional Neural Network for Kidney Tumor Segmentation in CT Images. *IEEE Trans. Image Process.* **2019**, *28*, 4060–4074. [CrossRef] [PubMed]
52. Pang, S.; Du, A.; Orgun, M. A.; Yu, Z.; Wang, Y.; Wang, Y.; Liu, G. CTumorGAN: A Unified Framework for Automatic Computed Tomography Tumor Segmentation. *Eur. J. Nucl. Med. Mol. Imaging* **2020**, *47*, 2248–2268. [CrossRef] [PubMed]
53. Shen, Z.; Yang, H.; Zhang, Z.; Zheng, S. Automated Kidney Tumor Segmentation with Convolution and Transformer Network. Available online: <https://openreview.net/forum?id=votlNyy36u> (accessed on 10 October 2021).
54. Yang, G.; Li, G.; Pan, T.; Kong, Y. Automatic Segmentation of Kidney and Renal Tumor in CT Images Based on 3D Fully Convolutional Neural Network with Pyramid Pooling Module. *Int. Conf. Pattern Recognit.* **2018**, *31571001*, 3790–3795. [CrossRef]
55. Heo, J. Automatic Segmentation in Abdominal CT Imaging for the KiTS21 Challenge. Available online: <https://openreview.net/forum?id=n6DR2TdGLa> (accessed on 20 October 2021).
56. Lund, C.B.; van der Velden, B.H.M. Leveraging Clinical Characteristics for Improved Deep Learning-Based Kidney Tumor Segmentation on CT. *arXiv* **2021**, arXiv:2109.05816.

57. Lin, C.; Fu, R.; Zheng, S. Kidney and Kidney Tumor Segmentation Using a Two-Stage Cascade Framework. Available online: <https://openreview.net/forum?id=TDOSVEQ8mdO> (accessed on 20 October 2021).
58. da Cruz, L.B.; Araújo, J.D.L.; Ferreira, J.L.; Diniz, J.O.B.; Silva, A.C.; de Almeida, J.D.S.; de Paiva, A.C.; Gattass, M. Kidney Segmentation from Computed Tomography Images Using Deep Neural Network. *Comput. Biol. Med.* **2020**, *123*, 103906. [CrossRef]
59. Zhang, Y.; Wang, Y.; Hou, F.; Yang, J.; Xiong, G.; Tian, J.; Zhong, C. Cascaded Volumetric Convolutional Network for Kidney Tumor Segmentation from CT Volumes. *arXiv* **2019**, arXiv:1910.02235. <https://doi.org/10.24926/548719.004>.
60. Hou, X.; Xie, C.; Li, F.; Wang, J.; Lv, C.; Xie, G.; Nan, Y. A Triple-Stage Self-Guided Network for Kidney Tumor Segmentation. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; pp. 341–344. [CrossRef]
61. Hatamizadeh, A.; Terzopoulos, D.; Myronenko, A. Edge-Gated CNNs for Volumetric Semantic Segmentation of Medical Images. *arXiv* **2020**, arXiv:2002.04207.
62. Santini, G.; Moreau, N.; Rubeaux, M. Kidney Tumor Segmentation Using an Ensembling Multi-Stage Deep Learning Approach. A Contribution to the KiTS19 Challenge. *arXiv* **2019**, arXiv:1909.00735.
63. Chen, Z.; Liu, H. 2.5D Cascaded Semantic Segmentation for Kidney Tumor Cyst. Available online: https://openreview.net/forum?id=d5WM4_asJCI (accessed on 20 October 2021).
64. Wei, H.; Wang, Q.; Zhao, W. Two-Phase Framework for Automatic Kidney and Kidney Tumor Segmentation. Available online: <https://kits.lib.umn.edu/two-phase-framework-for-automatic-kidney-and-kidney-tumor-segmentation/> (accessed on 15 October 2021).
65. He, T.; Zhang, Z.; Pei, C.; Huang, L. A Two-Stage Cascaded Deep Neural Network with Multi-Decoding Paths for Kidney Tumor Segmentation. Available online: <https://openreview.net/forum?id=c7kCK-E-B1> (accessed on 10 October 2021).
66. Lv, Y.; Wang, J. Three Uses of One Neural Network: Automatic Segmentation of Kidney Tumor and Cysts Based on 3D U-Net. Available online: https://openreview.net/forum?id=UzoGQ8fO_8f (accessed on 15 October 2021).
67. Li, D.; Chen, Z.; Hassan, H.; Xie, W.; Huang, B. A Cascaded 3D Segmentation Model for Renal Enhanced CT Images. Available online: <https://openreview.net/forum?id=dKvuhx2UPO3> (accessed on 10 October 2021).
68. Xiao, C.; Hassan, H.; Huang, B. Contrast-Enhanced CT Renal Tumor Segmentation. Available online: <https://openreview.net/forum?id=-QutS3TdRu-> (accessed on 20 October 2021).
69. Wen, J.; Li, Z.; Shen, Z.; Zheng, Y. Squeeze-and-Excitation Encoder-Decoder Network for Kidney and Kidney Tumor Segmentation in CT Images. Available online: <https://openreview.net/forum?id=uC-GL3IG8wn> (accessed on 10 October 2021).
70. Qayyum, A.; Lalande, A.; Meriaudeau, F. Automatic Segmentation of Tumors and Affected Organs in the Abdomen Using a 3D Hybrid Model for Computed Tomography Imaging. *Comput. Biol. Med.* **2020**, *127*, 104097. 2020.104097. [CrossRef] [PubMed]
71. Cheng, J.; Liu, J.; Liu, L.; Pan, Y.; Wang, J. *A Double Cascaded Framework Based on 3D SEAU-Net for Kidney and Kidney Tumor Segmentation*; University of Minnesota Libraries Publishing: Mankato, MN, USA, 2019; pp. 1–6.
72. 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. Kidney Tumor Segmentation from Computed Tomography Images Using DeepLabv3+ 2.5D Model. *Expert Syst. Appl.* **2021**, *192*, 116270. [CrossRef]
73. Zhao, W.; Zeng, Z. Multi Scale Supervised 3D U-Net for Kidney and Tumor Segmentation. *arXiv* **2019**, arXiv:1908.03204.
74. Meyer, J.G. Deep Learning Neural Network Tools for Proteomics. *Cell Rep. Methods* **2021**, *1*, 100003. [CrossRef]
75. Fukushima, K. Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biol. Cybern.* **1980**, *36*, 193–202. [CrossRef]
76. Baumgartl, H.; Tomas, J.; Buettner, R.; Merkel, M. A Deep Learning-Based Model for Defect Detection in Laser-Powder Bed Fusion Using in-Situ Thermographic Monitoring. *Prog. Addit. Manuf.* **2020**, *5*, 277–285. [CrossRef]
77. Buettner, R.; Baumgartl, H. A Highly Effective Deep Learning Based Escape Route Recognition Module for Autonomous Robots in Crisis and Emergency Situations. *Proc. Annu. Hawaii Int. Conf. Syst. Sci.* **2019**, *2019*, 659–666. [CrossRef]
78. Dalal, N.; Triggs, B. Histograms of Oriented Gradients for Human Detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; pp. 886–893. [CrossRef]
79. Bay, H.; Tuytelaars, T.; Van Gool, L. SURF: Speeded Up Robust Features. *Comput. Vis. Image Underst.* **2006**, *110*, 404–417.
80. Fazl-Ersi, E.; Tsotsos, J. K. Histogram of Oriented Uniform Patterns for Robust Place Recognition and Categorization. *Int. J. Rob. Res.* **2012**, *31*, 468–483. [CrossRef]
81. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-Based Learning Applied to Document Recognition. *Proc. IEEE* **1998**, *86*, 2278–2323. [CrossRef]
82. Koushik, J. Understanding Convolutional Neural Networks. *arXiv* **2016**, arXiv:1605.09081.
83. Ronneberger, O.; Fischer, P.; Brox, T. *U-Net: Convolutional Networks for Biomedical Image Segmentation* BT-*Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
84. Štern, D.; Bischof, H.; Urschler, M. *Regressing Heatmaps for Multiple Landmark Localization Using CNNs* Automatic Age Estimation from Skeletal and Dental MRI Data Using Machine Learning View Project SEE PROFILE *Regressing Heatmaps for Multiple Landmark Localization Using CNNs*; Springer International Publishing: Berlin/Heidelberg, Germany, 2016.

85. Mkrtchyan, K.; Singh, D.; Liu, M.; Reddy, V.; Roy-Chowdhury, A.; Gopi, M. Efficient Cell Segmentation and Tracking of Developing Plant Meristem. In Proceedings of the 2011 18th IEEE International Conference on Image Processing, Brussels, Belgium, 11–14 September 2011; pp. 2165–2168. [\[CrossRef\]](#)
86. Chang, C.S.; Ding, J.J.; Chen, P.H.; Wu, Y.F.; Lin, S.J. 3-D Cell Segmentation by Improved V-Net Architecture Using Edge and Boundary Labels. In Proceedings of the 2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP), Weihai, China, 28–30 September 2019; pp. 435–439. [\[CrossRef\]](#)
87. Shanthi, T.; Sabeenian, R. S. Modified Alexnet Architecture for Classification of Diabetic Retinopathy Images. *Comput. Electr. Eng.* **2019**, *76*, 56–64. [\[CrossRef\]](#)
88. Jiang, H.; Diao, Z.; Yao, Y.-D. Deep Learning Techniques for Tumor Segmentation: A Review. *J. Supercomput.* **2021**. [\[CrossRef\]](#)
89. Shen, H.; Wang, R.; Zhang, J.; McKenna, S. J. Boundary-Aware Fully Convolutional Network for Brain Tumor Segmentation. *Lect. Notes Comput. Sci.* **2017**, *10434*, 433–441. [\[CrossRef\]](#)
90. Havaei, M.; Davy, A.; Warde-Farley, D.; Biard, A.; Courville, A.; Bengio, Y.; Pal, C.; Jodoin, P. M.; Larochelle, H. Brain Tumor Segmentation with Deep Neural Networks. *Med. Image Anal.* **2017**, *35*, 18–31. [\[CrossRef\]](#)
91. Wang, G.; Li, W.; Ourselin, S.; Vercauteren, T. Automatic Brain Tumor Segmentation Based on Cascaded Convolutional Neural Networks with Uncertainty Estimation. *Front. Comput. Neurosci.* **2019**, *13*, 56. [\[CrossRef\]](#) [\[PubMed\]](#)
92. Wang, Z.; Wang, E.; Zhu, Y. *Image Segmentation Evaluation: A Survey of Methods*; Springer: Berlin/Heidelberg, Germany, 2020; Volume 53. [\[CrossRef\]](#)
93. Di Leo, G.; Di Terlizzi, F.; Flor, N.; Morganti, A.; Sardanelli, F. Measurement of Renal Volume Using Respiratory-Gated MRI in Subjects without Known Kidney Disease: Intraobserver, Interobserver, and Interstudy Reproducibility. *Eur. J. Radiol.* **2011**, *80*, e212–e216. [\[CrossRef\]](#) [\[PubMed\]](#)
94. King, B.F.; Reed, J. E.; Bergstralh, E.J.; Sheedy, P.F.; Torres, V.E. Quantification and Longitudinal Trends of Kidney, Renal Cyst, and Renal Parenchyma Volumes in Autosomal Dominant Polycystic Kidney Disease. *J. Am. Soc. Nephrol.* **2000**, *11*, 1505–1511. [\[CrossRef\]](#) [\[PubMed\]](#)
95. Taha, A.A.; Hanbury, A. Metrics for Evaluating 3D Medical Image Segmentation: Analysis, Selection, and Tool. *BMC Med. Imaging* **2015**, *15*, 1–28. [\[CrossRef\]](#) [\[PubMed\]](#)
96. Dogra, D.P.; Majumdar, A.K.; Sural, S. Evaluation of Segmentation Techniques Using Region Area and Boundary Matching Information. *J. Vis. Commun. Image Represent.* **2012**, *23*, 150–160. [\[CrossRef\]](#)
97. Karimi, D.; Salcudean, S.E. Reducing the Hausdorff Distance in Medical Image Segmentation with Convolutional Neural Networks. *IEEE Trans. Med. Imaging* **2020**, *39*, 499–513. [\[CrossRef\]](#) [\[PubMed\]](#)
98. Kirasich, K.; Smith, T.; Sadler, B. Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets. *Data Sci. Rev.* **2018**, *1*, 9.
99. Kaur, R.; Juneja, M.; Mandal, A.K. A Hybrid Edge-Based Technique for Segmentation of Renal Lesions in CT Images. *Multimed. Tools Appl.* **2019**, *78*, 12917–12937. [\[CrossRef\]](#)
100. Dill, V.; Franco, A.R.; Pinho, M.S. Automated Methods for Hippocampus Segmentation: The Evolution and a Review of the State of the Art. *Neuroinformatics* **2015**, *13*, 133–150. [\[CrossRef\]](#)
101. Yu, J.; Xu, J.; Chen, Y.; Li, W.; Wang, Q.; Yoo, B.; Han, J.-J. Learning Generalized Intersection Over Union for Dense Pixelwise Prediction. *PMLR* **2021**, *139*, 12198–12207.
102. Wu, H.; Yang, S.; Huang, Z.; He, J.; Wang, X. Type 2 Diabetes Mellitus Prediction Model Based on Data Mining. *Inform. Med. Unlocked* **2018**, *10*, 100–107. [\[CrossRef\]](#)
103. Saito, T.; Rehmsmeier, M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS ONE* **2015**, *10*, e0118432. [\[CrossRef\]](#) [\[PubMed\]](#)
104. Tahir, B.; Iqbal, S.; Usman Ghani Khan, M.; Saba, T.; Mehmood, Z.; Anjum, A.; Mahmood, T. Feature Enhancement Framework for Brain Tumor Segmentation and Classification. *Microsc. Res. Tech.* **2019**, *82*, 803–811. [\[CrossRef\]](#) [\[PubMed\]](#)
105. Sled, J.G.; Zijdenbos, A.P.; Evans, A.C. A Nonparametric Method for Automatic Correction of Intensity Nonuniformity in Mri Data. *IEEE Trans. Med. Imaging* **1998**, *17*, 87–97. [\[CrossRef\]](#) [\[PubMed\]](#)
106. Tustison, N.J.; Avants, B.B.; Cook, P.A.; Zheng, Y.; Egan, A.; Yushkevich, P.A.; Gee, J.C. N4ITK: Improved N3 Bias Correction. *IEEE Trans. Med. Imaging* **2010**, *29*, 1310–1320. [\[CrossRef\]](#)
107. Nyúl, L.G.; Udupa, J.K.; Zhang, X. New Variants of a Method of MRI Scale Standardization. *IEEE Trans. Med. Imaging* **2000**, *19*, 143–150. [\[CrossRef\]](#) [\[PubMed\]](#)
108. Kamnitsas, K.; Ledig, C.; Newcombe, V.F.J.; Simpson, J.P.; Kane, A.D.; Menon, D.K.; Rueckert, D.; Glocker, B. Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation. *Med. Image Anal.* **2017**, *36*, 61–78. [\[CrossRef\]](#)
109. Gordillo, N.; Montseny, E.; Sobrevilla, P. State of the Art Survey on MRI Brain Tumor Segmentation. *Magn. Reson. Imaging* **2013**, *31*, 1426–1438. [\[CrossRef\]](#)
110. Dong, H.; Yang, G.; Liu, F.; Mo, Y.; Guo, Y. Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks. *Commun. Comput. Inf. Sci.* **2017**, *723*, 506–517. [\[CrossRef\]](#)
111. Torres, H.R.; Queirós, S.; Morais, P.; Oliveira, B.; Fonseca, J.C.; Vilaça, J.L. Kidney Segmentation in Ultrasound, Magnetic Resonance and Computed Tomography Images: A Systematic Review. *Comput. Methods Programs Biomed.* **2018**, *157*, 49–67. [\[CrossRef\]](#)

112. Udupa, J.K.; Leblanc, V.R.; Schmidt, H.; Imielinska, C.; Saha, P.K. A Methodology for Evaluating Image Segmentation Algorithms Medical Image Processing Group—Department of Radiology—University of Pennsylvania. *Image Processing. Int. Soc. Opt. Photonics* **2002**, *4684*, 266–277.
113. Fasihi, M.S.; Mikhael, W.B. Overview of Current Biomedical Image Segmentation Methods. In Proceedings of the 2016 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 15–17 December 2016; pp. 803–808. [\[CrossRef\]](#)
114. Ramkumar, A.; Dolz, J.; Kirisli, H.A.; Adebahr, S.; Schimek-Jasch, T.; Nestle, U.; Massoptier, L.; Varga, E.; Stappers, P.J.; Niessen, W.J.; Song, Y. User Interaction in Semi-Automatic Segmentation of Organs at Risk: A Case Study in Radiotherapy. *J. Digit. Imaging* **2016**, *29*, 264–277. [\[CrossRef\]](#) [\[PubMed\]](#)
115. Bae, K.T.; Shim, H.; Tao, C.; Chang, S.; Wang, J.H.; Boudreau, R.; Kwok, C.K. Intra- and Inter-Observer Reproducibility of Volume Measurement of Knee Cartilage Segmented from the OAI MR Image Set Using a Novel Semi-Automated Segmentation Method. *Osteoarthr. Cartil.* **2009**, *17*, 1589–1597. [\[CrossRef\]](#) [\[PubMed\]](#)
116. Gaonkar, B.; Macyszyn, L.; Bilello, M.; Sadaghiani, M.S.; Akbari, H.; Atthiah, M.A.; Ali, Z.S.; Da, X.; Zhan, Y.; Rourke, D.O.; Grady, S.M.; Davatzikos, C. Automated Tumor Volumetry Using Computer-Aided Image Segmentation. *Acad. Radiol.* **2015**, *22*, 653–661. [\[CrossRef\]](#)
117. Graber, C.; Tsai, G.; Firman, M.; Brostow, G.; Schwing, A. Panoptic Segmentation Forecasting. *arXiv* **2021**, arXiv:2104.03962.
118. Thoma, M. A Survey of Semantic Segmentation. *arXiv* **2016**, arXiv:1602.06541.
119. Liu, X.; Deng, Z.; Yang, Y. Recent Progress in Semantic Image Segmentation. *Artif. Intell. Rev.* **2019**, *52*, 1089–1106. [\[CrossRef\]](#)
120. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Garcia-Rodriguez, J. A Review on Deep Learning Techniques Applied to Semantic Segmentation. *arXiv* **2017**, arXiv:1704.06857.
121. Bou, A. Deep Learning Models for Semantic Segmentation of Mammography Screenings. 2019. Available online: <https://www.diva-portal.org/smash/record.jsf?pid=diva2:1380578> (accessed on 10 October 2021).
122. Menze, B.H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans. Med. Imaging* **2015**, *34*, 1993–2024. [\[CrossRef\]](#)
123. Arel, I.; Rose, D.; Karnowski, T. Deep Machine Learning—A New Frontier in Artificial Intelligence Research. *IEEE Comput. Intell. Mag.* **2010**, *5*, 13–18. [\[CrossRef\]](#)
124. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [\[CrossRef\]](#)
125. Collobert, R.; Weston, J. A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning. In Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland, 5–9 July 2008; pp. 160–167. [\[CrossRef\]](#)
126. Lee, H.; Yan, L.; Pham, P.; Ng, A.Y. Unsupervised Feature Learning for Audio Classification Using Convolutional Deep Belief Networks. *Adv. Neural Inf. Process. Syst. Proc.* **2009**, 1096–1104. Available online: <https://proceedings.neurips.cc/paper/2009/hash/a113c1ecd3cace2237256f4c712f61b5-Abstract.html> (accessed on 10 October 2021).
127. Graham-Knight, J.B.; Djavadifar, A.; Lasserre, D.P.; Najjaran, H. Applying NnU-Net to the KiTS19 Grand Challenge. *Univ. Minn. Libr.* **2019**, 1–7. [\[CrossRef\]](#)
128. Rao, P.K.; Chatterjee, S. WP-UNet: Weight Pruning U-Net with Depthwise Separable Convolutions for Semantic Segmentation of Kidney Tumors. *ResearchSquare* **2021**. [\[CrossRef\]](#)
129. Sabarinathan, D.; Parisa Beham, M.; Mansoor Roomi, S.M.M. Hyper Vision Net: Kidney Tumor Segmentation Using Coordinate Convolutional Layer and Attention Unit. *Commun. Comput. Inf. Sci.* **2020**, *1249*, 609–618. [\[CrossRef\]](#)
130. Yuan, Y. Automatic Kidney and Tumor Segmentation with Hybrid Hierarchical Networks. *Comput. Biol. Med.* **2019**. [\[CrossRef\]](#)
131. Sharma, R.; Halarnkar, P.; Choudhari, K. Kidney and Tumor Segmentation Using U-Net Deep Learning Model. *SSRN Electron. J.* **2020**, 2020. [\[CrossRef\]](#)
132. Li, C.; Chen, W.; Tan, Y. Point-Sampling method Based on 3DU-Net Architecture to Reduce the Influence of False Positive and Solve Boundary Blur Problem in 3DCT Image Segmentation. *Appl. Sci.* **2020**, *10*, 6838. [\[CrossRef\]](#)
133. Akram, Z. Cancerous Tumor Segmentation of Kidney Images and Prediction of Tumor Using Medical Image Segmentation and Deep Learning Techniques. *Clin. Oncol.* **2021**, *4*, 1–9.
134. Daza, L.; Gómez, C.; Arbeláez, P. Semantic Segmentation of Kidney Tumor Using Convolutional Neural Networks. *Univ. Minn. Libr.* **2019**. [\[CrossRef\]](#)
135. Shen, C.; Wang, C.; Oda, M.; Mori, K. Coarse-to-Fine Kidney and Tumor Segmentation with Fully Convolutional Networks. *arXiv* **2019**, arXiv:1908.11064v1.
136. Vu, M.H.; Grimmer, G.; Simkó, A.; Nyholm, T.; Löfstedt, T. Localization Network and End-to-End Cascaded U-Nets for Kidney Tumor Segmentation. *Univ. Minn. Libr.* **2019**. [\[CrossRef\]](#)
137. Lv, Y.; Wang, J. Kidney Tumor Segmentation Based on U-Net and V-Net with Double Loss Function Training. *Univ. Minn. Libr.* **2019**, 1–4. [\[CrossRef\]](#)
138. Xia, K.; Yin, H.; Zhang, Y. Deep Semantic Segmentation of Kidney and Space-Occupying Lesion Area Based on SCNN and ResNet Models Combined with SIFT-Flow Algorithm. *J. Med. Syst.* **2019**, *43*, 1–12. [\[CrossRef\]](#)

139. Bakas, S.; Reyes, M.; Jakab, A.; Bauer, S.; Rempfler, M.; Crimi, A.; Shinohara, R. T.; Berger, C.; Ha, S. M.; Rozycki, M.; et al. Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. *arXiv* **2018**, arXiv:1811.02629.
140. Shorten, C.; Khoshgoftaar, T.M. A Survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 1–48. [[CrossRef](#)]
141. Pavlov, S.; Artemov, A.; Sharaev, M.; Bernstein, A.; Burnaev, E. Weakly Supervised Fine Tuning Approach for Brain Tumor Segmentation Problem. In Proceedings of the 2019 18th IEEE International Conference On Machine Learning and Applications (ICMLA), Boca Raton, FL, USA, 16–19 December 2019; pp. 1600–1605. [[CrossRef](#)]
142. Sarker, I.H. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Comput. Sci.* **2021**, *2*, 1–20. [[CrossRef](#)]
143. Zikic, D.; Ioannou, Y.; Brown, M.; Criminisi, A. Segmentation of Brain Tumor Tissues with Convolutional Neural Networks. *Proc. MICCAI-BRATS* **2014**, *2014*, 36–39.
144. Wang, L.; Wang, S.; Chen, R.; Qu, X.; Chen, Y.; Huang, S.; Liu, C. Nested Dilation Networks for Brain Tumor Segmentation Based on Magnetic Resonance Imaging. *Front. Neurosci.* **2019**, *13*, 1–14. [[CrossRef](#)] [[PubMed](#)]

2.2 Conclusion

In conclusion, the paper "Kidney Tumor Semantic Segmentation Using Deep Learning: A Survey of State-of-the-Art" provides a thorough overview of the latest advancements in deep learning techniques applied to kidney tumor segmentation. It critically analyzes various state-of-the-art CNN architectures, highlighting their performance, strengths, and limitations. The paper also examines the evaluation metrics commonly used to assess these models, offering insights into their applicability in real-world scenarios. By exploring key factors that influence model performance, the paper identifies opportunities for future research to improve the accuracy and effectiveness of kidney tumor segmentation systems. Overall, this survey serves as a valuable resource for understanding the current landscape and guiding future innovations in the field.

Chapter 3

EfficientNet family U-Net models for deep learning semantic segmentation of kidney tumors on CT images

3.1 Introduction

The paper "EfficientNet Family U-Net Models for Deep Learning Semantic Segmentation of Kidney Tumors on CT Images" explores the use of EfficientNet-based U-Net architectures for segmenting kidney tumors in CT scans. By leveraging the efficiency and scalability of the EfficientNet family, combined with the U-Net's strengths in medical image segmentation, the models achieve high accuracy while maintaining computational efficiency. The study highlights the models' ability to balance performance with resource usage, making them suitable for clinical applications where accurate and fast kidney tumor segmentation is critical. Overall, the paper demonstrates the effectiveness of these models in improving tumor detection and segmentation.



OPEN ACCESS

EDITED BY

Nicola Strisciuglio,
University of Twente, Netherlands

REVIEWED BY

Amir Faisal,
Sumatra Institute of Technology, Indonesia
Virginia Riego Del Castillo,
University of León, Spain
Lidia Talavera-Martinez,
University of the Balearic Islands, Spain

*CORRESPONDENCE

Serestina Viriri
✉ viriris@ukzn.ac.za

RECEIVED 06 June 2023

ACCEPTED 07 August 2023

PUBLISHED 07 September 2023

CITATION

Abdelrahman A and Viriri S (2023) EfficientNet family U-Net models for deep learning semantic segmentation of kidney tumors on CT images. *Front. Comput. Sci.* 5:1235622. doi: 10.3389/fcomp.2023.1235622

COPYRIGHT

© 2023 Abdelrahman and Viriri. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

EfficientNet family U-Net models for deep learning semantic segmentation of kidney tumors on CT images

Abubaker Abdelrahman and Serestina Viriri *

School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, South Africa

Introduction: Kidney tumors are common cancer in advanced age, and providing early detection is crucial. Medical imaging and deep learning methods are increasingly attractive for identifying and segmenting kidney tumors. Convolutional neural networks have successfully classified and segmented images, enabling clinicians to recognize and segment tumors effectively. CT scans of kidneys aid in tumor assessment and morphology study, using semantic segmentation techniques for pixel-level identification of kidney and surrounding anatomy. Accurate diagnostic procedures are crucial for early detection of kidney cancer.

Methods: This paper proposes an EfficientNet model for complex segmentation by linking the encoder stage EfficientNet with U-Net. This model represents a more successful system with improved encoder and decoder features. The Intersection over Union (IoU) metric quantifies model performance.

Results and Discussion: The EfficientNet models showed high IoU_Scores for background, kidney, and tumor segmentation, with mean IoU_Scores ranging from 0.976 for B0 to 0.980 for B4. B7 received the highest IoU_Score for segmenting kidneys, while B4 received the highest for segmenting tumors. The study utilizes the KiTS19 dataset for contrast-enhanced CT images. Using Semantic segmentation for EfficientNet Family U-Net Models, our method proved even more reliable and will aid doctors in accurate tumor detection and image classification for early diagnosis.

KEYWORDS

kidney tumor, EfficientNet, deep learning, U-Net, semantic segmentation

1. Introduction

Many factors contribute to the progressive rise in cancer incidence in humans, where effective disease prevention requires early identification and treatment. One of the essential organs in human physiology is the kidney. Kidney cancer requires a precise diagnosis and structured care (Rajinikanth et al., 2023). One of the top 10 cancers that affect both men and women is kidney cancer. The lifetime risk of developing kidney cancer is around 1 in 75 (1.34%). Kidney cancer ranks ninth in men and fourteenth in women. If kidney cancer is detected and treated in its early stages, afflicted individuals have a far higher chance of being cured (Hsiao et al., 2022). The human body has two kidneys and bean-shaped organs on either side of the spine. The kidney plays a crucial role in maintaining body fluid and solute balance through the excretion and filtration of waste products.

It also secretes a variety of hormones and aids in blood pressure regulation. A kidney tumor develops when kidney cells stop functioning and begin to grow quickly. Renal cell carcinoma, which develops from kidney cells, can spread slowly or aggressively. Renal cell cancer frequently manifests as a single tumor. However, any kidney may develop a variety of malignancies (Geethanjali and Dinesh, 2021; Abdelrahman and Viriri, 2022).

Kidney cancer diagnosis usually does not begin with palpation. In most cases, kidney cancer is asymptomatic in its early stages and is often detected incidentally on imaging studies performed for unrelated reasons. Palpation of the kidneys during a physical exam may detect an enlarged kidney, but it is unreliable for detecting kidney cancer (Abdelrahman and Viriri, 2022; Bapir et al., 2022). Once kidney cancer is suspected based on imaging studies, a diagnosis is typically confirmed with a biopsy or surgical tumor resection. A biopsy involves taking a small sample of the tumor tissue for examination under a microscope, while surgical resection involves removing the entire tumor for examination; additional tests may also be performed to determine the extent of cancer and guide treatment decisions. These may include blood tests to assess kidney function and identify any markers of cancer, such as elevated levels of certain proteins, and imaging studies, such as CT or MRI scans, to evaluate the size and location of the tumor and identify any metastases. Follow-up tests include urine analysis, general examinations, lymph node palpations every 2 years, and cystoscopies every 10 years to check for the absence of renal cell carcinoma recurrence (Beisland et al., 2006; Üreyen et al., 2015; Abdelrahman and Viriri, 2022; Bapir et al., 2022).

Manual segmentation methods have been used extensively to study renal carcinoma. Manual kidney segmentation manually outlines the kidney region in medical images by outlining the kidney region and identifying the location and extent of tumors. This is an important step in many clinical applications. Manual kidney segmentation can be time-consuming and subject to inter-observer variability, meaning that different radiologists may segment the kidneys differently. However, it is still widely used as a gold standard for evaluating the accuracy of automated segmentation algorithms. By comparing the results of a machine learning algorithm with manual segmentation, researchers can determine the algorithm's performance and identify areas for improvement (Torres et al., 2018; Abdelrahman and Viriri, 2022). As a result, renal segmentation is an important stage in computer-aided analysis and is particularly important for early cancer detection. In addition to carcinoma, the exact segmentation provides structural evidence of variations in the size and form of the kidneys and tumors, which professionals may utilize to examine significant disorders. Because of the importance of this field to medicine, several researchers choose to employ image processing, machine learning, and deep learning methodologies to widen computational tools to aid experts in the interpretation of clinical images (Hatipoglu and Bilgin, 2017; Nazari et al., 2021; Abdelrahman and Viriri, 2022).

Deep learning techniques have recently shown promising results. It took a lot of time and effort to build earlier attempts to categorize kidney cancer histology images using manual feature extraction techniques and conventional machine learning algorithms. On the other hand, deep learning techniques automate

this procedure (Gurcan et al., 2009). Hence, early identification is essential to guarantee early diagnosis in kidney cancer candidates, boost treatment effectiveness, and decrease death rates. The medical industry has profited from the development and progress of machine learning, yet there is still much opportunity for improvement.

1.1. Research problem

Given the shortage of available medical experts (Kandel and Castelli, 2020; Munien and Viriri, 2021), the potential for inter-observer variability, and the laborious process of making a diagnosis of kidney cancer, all support the need for an automated system to classify kidney cancer histopathology images accurately. By using advanced computer algorithms and machine learning techniques, such a system could potentially improve the speed and accuracy of cancer diagnosis, leading to earlier detection and better patient outcomes. However, it is important to note that any AI-based medical system must be thoroughly validated and tested to ensure its safety and effectiveness before use in clinical practice. Previous approaches have shown promise in addressing this issue. The research explores eight lightweight EfficientNet family architectures, focusing on optimizing resources while maintaining high accuracies, making them valuable tools for resource optimization tasks. They may achieve comparable results to state-of-the-art approaches while consuming less space and training time. This research aims to answer the capability of EfficientNets to achieve similar results to state-of-the-art approaches for classifying kidney cancer CT images (Aresta et al., 2019; Tan and Le, 2019; Abdelrahman and Viriri, 2022). In the next part, we will explore a selection of scholarly papers authored by teachers, focusing specifically on the field of kidney and kidney tumor segmentation.

1.2. Literature review

Computer-aided diagnosis (CAD) systems assist healthcare professionals in making diagnoses by providing additional information and analysis based on medical imaging and data. They use machine learning, deep learning, and image processing to analyze and interpret patient data, emphasizing areas of concern and diagnoses. However, CAD systems are not meant to replace healthcare experts but rather to supplement their knowledge and aid decision-making. Medical practitioners must ultimately diagnose patients and choose appropriate therapy courses based on their clinical judgment and expertise (Ramadan, 2020; Munien and Viriri, 2021; Abdelrahman and Viriri, 2022; Tsuneki, 2022).

1.2.1. Modern approaches

Recent growth in processing power has led to advancements in deep learning-based technologies, particularly CNNs, in medical image processing. These technologies have been successful in segmentation detection, abnormality classification, and retrieval, resulting in the emergence of intriguing algorithms in this field

(Xie et al., 2020). In Ali et al. (2007), the authors use contour templates to describe the shapes of local objects, and Contour templates describe the shapes of local objects like organs, in medical images. These pre-defined shapes capture specific properties of different organs, making them flexible and efficient for segmenting organs without complex models or training data. They can be customized to capture specific features, making them ideal for accurate segmentation in medical imaging for diagnosis and treatment planning (Thong et al., 2018). 2D CNNs are used in medical imaging to segment CT images using two common deep learning segmentation strategies: end-to-end segmentation and cascade segmentation. End-to-end segmentation uses a single deep-learning model, making it simpler and more efficient. However, it may have reduced flexibility, operability, and interoperability, and may require more training data. Cascade segmentation divides the task into multiple stages, using a separate deep-learning model for each stage. Both approaches are widely used in medical image segmentation, depending on the task's requirements, image data complexity, and available computational resources (Xie et al., 2020). Deep Learning focuses on artificial neural networks with multiple layers, used in natural language processing, gaming, image and audio recognition, and assessing visual images (Valueva et al., 2020). CNNs replace matrix multiplication in image processing and recognition, focusing on pixel data (Kim, 2019). They are utilized in image recognition, recommendation systems, classification, segmentation, medical analysis, and more (Collobert and Weston, 2008; Tsantekidis et al., 2017; Avilov et al., 2020).

1.2.2. Convolutional neural network approach

Convolutional neural networks (CNNs) are highly effective learning algorithms for image segmentation, classification, detection, and retrieval tasks (Ciregan et al., 2012; Mayer et al., 2017). CNNs are efficient for image recognition due to their hierarchical representations and spatial structure, making them ideal for categorizing histopathology photos in general computer vision applications (Khan et al., 2020). CNNs have been employed in several studies to identify and categories kidney cancers. The segmentation of kidneys and tumors in abdominal 3D CT scans is the specific emphasis of this section's discussion of several deep-learning techniques for the semantic segmentation of medical images. In the image (G-A), the yellow color indicates the kidney. In the image (G-B), the yellow color indicates Tumor and the green color indicates Kidney. Figure 1 depicts the segmentation of kidneys and tumors, highlighting their distinct regions. Images were taken from KiTS 19 dataset.

1.2.2.1. One-stage methods

One-stage approaches are object identification models that accurately forecast bounding boxes and class probabilities for each item in an image in a single forward pass. Convolutional neural networks (CNNs) use one-stage techniques to extract information from input images, produce feature maps, and predict object bounding boxes and class probabilities. Researchers use deep learning architectures like U-Net, LinkNet-34, and RAU-Net, and improve performance by combining data from multiple scales or training segmentation networks with non-squared patches.

Moreover, some techniques use transformers to record distant relationships for precise tumor segmentation. Some research examples and findings are shown below: The authors (Isensee and Maier-Hein, 2019) propose a "Reset U-Net"-based U-Net architecture for 3D medical image segmentation, addressing the vanishing gradient problem and improving performance. They suggest incorporating a "reset block" to facilitate gradient movement across the network. The authors Myronenko and Hatamizadeh (2019) suggest an automated strategy for segmenting kidney tumors in contrast-enhanced CT images based on deep learning.

Authors use a fully convolutional neural network to semantically segment kidney and tumor areas using KiTS19 challenge dataset contrast-enhanced CT images. The authors Geethanjali and Dinesh (2021) a novel Attention U-Net model used an attention mechanism and modified U-Net architecture to segment kidney tumors from CT data with 0.86 accuracies (Efremova et al., 2019). The research proposes an automated segmentation method for locating kidney and liver cancers in CT scans using a fully convolutional neural network architecture. The method accurately segments kidney and liver tumors, with an average Dice coefficient of 0.76 for kidney and 0.80 for liver tumors (Yang et al., 2018). The authors utilized a weighted loss function in network training to improve segmentation in CT images, potentially aiding radiologists in identifying and treating kidney and renal cancers. Methods include strategic sampling, weighted loss function, and transformers. The Table 1 displays the outcomes of the one-stage methods. All papers in one-stages use the KiTS19 dataset with CT images.

1.2.2.2. Two-stage methods

The two-stage approach in medical image segmentation addresses foreground/background imbalance, improving accuracy and efficiency by detecting the volume of interest (VOI) and segmenting target organs. Variations have strengths and weaknesses depending on application requirements.

Studies on two-stage kidney and tumor segmentation techniques in CT images have been conducted in several instances. da Cruz et al. (2020) created a technique that uses deep (CNN), and it managed to distinguish kidneys with up to 93.03% accuracy. Zhang et al. (2019) developed a cascaded two-stage kidney and tumor segmentation framework that uses a 3D fully convolutional network. Hou et al. (2020) provided a three-stage self-guided network for segmenting kidney tumors that finds the VOI using down-sampled CT images and extracts the kidney and tumor boundaries inside the VOI using the full-resolution net and tumor refine net from full-resolution CT images (Hatamizadeh et al., 2020). Module, which can be used with any general encoder-decoder architecture, improved the edge representations in learnt feature maps (Zhao et al., 2020) created the MSS U-Net, a multi-scale supervised 3D U-Net-based model for segmenting kidney and kidney tumors from CT images (Xie et al. (2020) presented a cascaded SE-ResNeXT U-Net, and (Chen and Liu, 2021) provided a method based on a multi-stage stepwise refinement strategy for segmenting kidney, tumor, and cyst in the abdomen-enhanced CT images. Finally, Wen et al. (2021) proposed the SeResUNet segmentation network, which leverages ResNet to deepen the encoder's network and speed convergence and is specifically

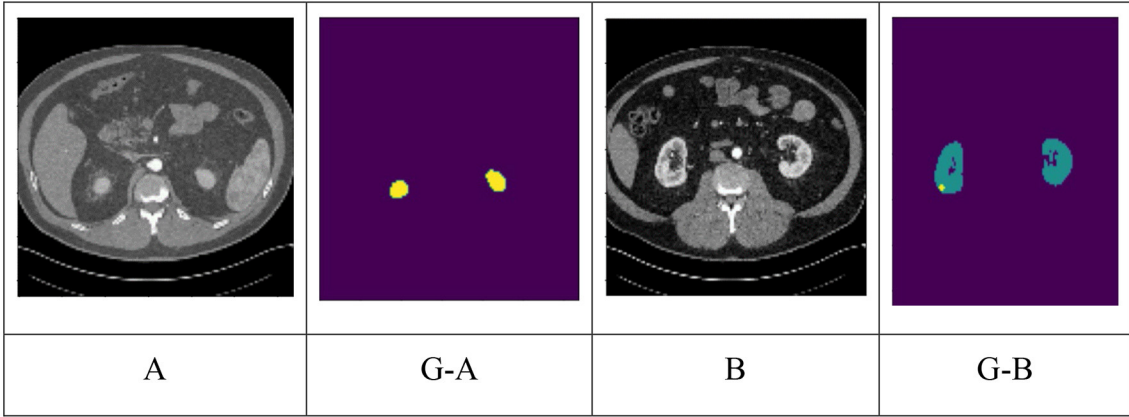


FIGURE 1
Different classes of kidney dataset (A) normal, (G-A) ground truth of (A). Yellow indicates kidney; (B) tumor; (G-B) ground truth of (B). Green indicates kidney and Yellow indicates tumor.

TABLE 1 Displays the outcomes of several one-stage methods.

Reference	Input	Regulation	Architecture	Loss	Optimizer	Kidneys dice	Tumor dice	Kidneys IoU scores	Tumor IoU scores
Isensee and Maier-Hein (2019)	3D		3D U-Net	Dice	SDG	0.973	0.850	-	-
Santini et al. (2019)	2.5D	BN	mult-istage-U-Net	Dice	Adam	0.96	0.74	-	-
Causey et al. (2021)	2D		Ensemble	Dice		0.947	0.778	-	-
Myronenko and Hatamizadeh (2019)	3D		Boundary-Aware Architecture	KD, TD, CD	Adam	0.974	0.810	-	-
Guo et al. (2021)	3D		RAU-Net	Dice SDG	Dice SDG	0.960	0.770	-	-
Efremova et al. (2019)	2D,3D		CNN	Dice		0.964	0.674	-	-
Geethanjali and Dinesh (2021)	3D		U-Net	IoU_Scores		-	-	0.957	0.939

designed to segment the kidney and tumor. Regarding kidney and tumor segmentation, two-stage approaches have generally shown encouraging results, but there is always potential for improvement, particularly for smaller organs and tumors. Table 2 displays the outcomes of the Two-Stage Methods. All papers in two-stages use the KiTS19 dataset with CT images except (Wen et al., 2021) use KiTS21 with CT images.

Two-stage medical image segmentation technique detects Voids and targets organs using deep convolutional neural networks, image processing methods, and architectures. Improvements are needed for smaller kidneys, tumors, and cysts. In kidney tumor semantic segmentation, one-stage methods offer a simpler and more efficient approach by employing a single neural network for end-to-end segmentation. This results in faster inference times and makes them well-suited for real-time applications. However, these methods may compromise accuracy and stress to capture fine details and handle complex structures. On the other hand, two-stage methods present a more sophisticated strategy with separate proposal generation and segmentation

refinement steps, leading to higher accuracy, robustness in handling challenging cases, and the ability to support hierarchical processing and feature reuse. Nevertheless, this increased accuracy comes at the cost of greater computational complexity, and they tend to have longer inference times, making them less ideal for real-time scenarios compared to their one-stage counterparts.

1.3. Research contributions

The study evaluated eight EfficientNets versions for kidney cancer CT image classification. The architecture effectively extracted and learned global image features, including tissue and nuclei organization. EfficientNet-B4 and EfficientNet-B7 models showed the best performance, with an accuracy mean IoU of 0.980. EfficientNet-B4 was superior for tumor detection and EfficientNet-B7 for kidney detection. The study highlights the potential of EfficientNets for kidney cancer and CT image classification due to their simplicity, reduced training time, and consistent accuracies.

TABLE 2 The outcomes of several two-stage methods.

Reference	Input	Regulation	Architecture	Loss	Optimizer	Kidneys dice	Tumor dice	Kidneys IoU scores	Tumor IoU scores
Zhang et al. (2019)	3D	BN	Cascaded	Dice	Adam	0.974	0.831	-	-
Hou et al. (2020)	3D		U-Net	Dice	Adam	0.967	0.845	-	-
Hatamizadeh et al. (2020)	3D		EG-CNN	Dice	Adam	0.965	0.835	-	-
Xie et al. (2020)	3D	BN	Cascaded	Dice	Adam	0.967	0.743	-	-
Wen et al. (2021)	3D		SeResUNet	Dice	Adam	0.916	0.542	-	-

1.4. Paper structure

This paper is organized into different sections. Part 2, Material and Method, the methods and techniques section, provides information on the framework used in this study. Part 3, Results and Discussion, is dedicated to presenting the results obtained during the research. Finally, Part 4, Conclusion, discusses the insights gained from this study and concludes the paper.

2. Material and method

The dataset used in this study is KiTS19, which stands for the Kidney Tumor Segmentation Challenge 2019. KiTS19 is a publicly available dataset of CT scans of the abdomen with kidney tumors (Heller et al., 2019). Evaluation Metrics: The evaluation metric used in this study is the Intersection over Union (IoU), also known as the Jaccard index. IoU is a common evaluation metric used for object detection and segmentation tasks.

2.1. Dataset

The test facts are gathered from KiTS19 (Heller et al., 2019). It is a dataset for kidney tumor segmentation that includes CT scans and corresponding annotations for training and testing machine learning models. It is a widely used benchmark dataset for evaluating the performance of medical image segmentation models on kidney tumors. Two hundred ten abdominal CT volumes comprise the testing and challenging training datasets. All slices are seen axially, and each volume's imaging and ground truth labels are supplied in Neuroimaging Informative Technology Informative (DICOM) format. This can be found on Kaggle at <https://www.kaggle.com/competitions/body-morphometry-kidney-and-tumor/data> (Who used the original data in Heller et al., 2019) with shape (number of slices, height, and breadth). The dataset comprises slices between 29 and 1059 pixels and grayscale images with a size of 512×512 pixels. The following Table 3, lists the characteristics of the experimental dataset. We use part of the data, 7,899 png images.

TABLE 3 Dataset properties used for the experimentation.

Properties	Values
Number of all image	7,899
Number of training images	5,841
Number of validation images	1,027
Number of testing images	1,031
Image format	Png
Modality	CT

2.2. Preprocessing

Preprocessing is indeed crucial for the classification of images. Images are typically stained with various dyes to highlight different structures and features, and they can vary significantly in terms of staining intensity, color, and texture. Therefore, preprocessing is necessary to enhance the images' quality and consistency and remove noise and artifacts that may interfere with the classification process (Aresta et al., 2019). They are rather large, while convolutional neural networks are typically designed to take in much smaller inputs. Therefore, the resolution of the images must be decreased so that the network can receive the input while maintaining the important features. In addition to the preprocessing steps mentioned earlier, another important step is to resize or down sample the CT images to a suitable size for processing by convolutional neural networks (CNNs). CNNs are typically designed to take in images of a fixed size, and the size of the input images can significantly impact the network's performance. Therefore, the resolution of the CT images must be decreased while preserving the important features and structures in the image. This can be achieved using down-sampling techniques, such as bilinear or nearest-neighbor interpolation, or more advanced techniques, such as wavelet-based or deep learning-based methods.

2.3. U-Net

U-Net is a convolutional neural network (CNN) architecture used for image segmentation tasks, which was proposed by Olaf Ronneberger, Philipp Fischer, and Thomas Brox in 2015. The network is named after its U-shape architecture, which consists

of a contracting path (encoder) and an expansive path (decoder). The contracting path consists of convolutional and max pooling layers, which reduce the spatial resolution of the input image while increasing the number of feature channels. This process enables the network to capture the high-level features of the image. The expansive path consists of transposed convolutional layers, which increase the spatial resolution of the feature maps while decreasing the number of feature channels. This path enables the network to perform pixel-wise classification by predicting a binary mask for each pixel in the input image. The U-Net architecture has become popular in the medical imaging community for its ability to segment organs and other structures from medical images accurately. It has also been used for other applications, such as the segmentation of objects in satellite images and the semantic segmentation of urban scenes (Ronneberger et al., 2015). The U-Net is a convolutional neural network for biomedical image segmentation. Its architecture, consisting of an encoder-decoder design, is highly stable and can achieve precise segmentation with fewer training images. The network comprises 3×3 convolutional layers, with a maximum of 2×2 following each pooling layer and the ReLU activation function. A 1×1 convolutional layer is attached at the end. The U-Net consists of path contracting and symmetric expanding, used for capturing and precise localization. The U-Net architecture depends heavily on data augmentation techniques. The network can segment a 512×512 image on a modern GPU in less than a second. The U-Net has been successfully applied to medical image segmentation, although 3D convolution is suggested for fully utilizing the spatial information of 3D images such as CT and MRI. Based on the 3D U-Net, Fabian made minor changes and achieved first place in several medical image segmentation contests, demonstrating that an optimized U-Net can outperform many innovative designs (Ronneberger et al., 2015; Payer et al., 2016).

2.4. EfficientNet

Convolutional neural network (CNN) researchers at Google AI unveiled the EfficientNet architecture in Figure 2. The primary objective of EfficientNet was to develop a high-accuracy and computationally efficient CNN that could be trained and used on various platforms, including mobile phones and other systems with limited resources. Compound scaling, which involves scaling each of the neural network's dimensions (such as depth, breadth, and resolution) in a balanced manner, is the foundation of the design of EfficientNet. This enables higher performance and efficiency than conventional scaling techniques, which often concentrate on just one or two network dimensions. EfficientNet uses a cutting-edge "compound coefficient optimization" technique to balance depth, breadth, and resolution scaling. This technique is one of the main inventions of EfficientNet. This enables higher accuracy levels while making better use of computer resources. On a variety of computer vision tasks, such as image classification, object recognition, and semantic segmentation, EfficientNet has produced state-of-the-art results. For instance, EfficientNet outperformed all other models in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2019 with a top-1 accuracy of 88.4% and a top-5 accuracy of 97.4%.

EfficientNet has been demonstrated to be extremely resource-efficient and have great accuracy. For instance, EfficientNet reduced parameter count by ten times and inference time by 2.8 times in one research compared to the prior state-of-the-art model on a mobile device. In both business and academics, EfficientNet is widely utilized for various purposes, including augmented reality, medical imaging, and autonomous driving. Researchers and developers may experiment with and modify the architecture because it is open-source (Tan and Le, 2019, 2021).

EfficientNet is a powerful convolutional neural network (CNN) architecture that has shown impressive results on a wide range of computer vision tasks. One such task is medical image analysis, where EfficientNet has been used to analyze medical images for various purposes, including cancer detection.

2.5. Evaluation metrics

The study used two evaluation metrics for evaluating image segmentation performance: IoU (Intersection over Union) and Dice similarity coefficient (DSC). IoU quantifies the similarity between predicted and ground truth masks by calculating the ratio of common pixels. DSC, also known as F1 score or Sørensen-Dice index, is a balanced measure of segmentation accuracy. Both metrics are appropriate for evaluating segmentation results, but the IoU is more sensitive to under- and over-segmentation errors. Despite this, DSC is the most commonly used metric in scientific publications for medical image segmentation evaluations. These metrics are calculated using the formulae shown below.

$$\text{IoU} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Positive} + \text{False Negative})} \quad (\text{Pandey et al., 2020}) \quad (1)$$

$$\text{DSC} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (\text{Pandey et al., 2020}) \quad (2)$$

2.6. Methods and techniques

The methodology presented in the study aims to segment tumors and kidneys on CT slices using the KiTS19 database (Heller et al., 2019). The method consists of two main stages: data preparation and segmentation using an EfficientNet U-Net model. In the data preparation stage, the CT volumes are first scaled and normalized to improve the consistency and quality of the data. Then, the volumes are transferred from DICOM to png format for further processing. The segmentation stage involves using an EfficientNet U-Net model to segment the kidneys and the tumor concurrently. The U-Net model is a type of convolutional neural network that is widely used for image segmentation tasks. The EfficientNet variant of the U-Net model is a more efficient and accurate version of the standard U-Net, which uses fewer parameters and achieves better results. Preprocessing is also essential for the methodology presented in the study involves data preparation and segmentation using an EfficientNet U-Net model.

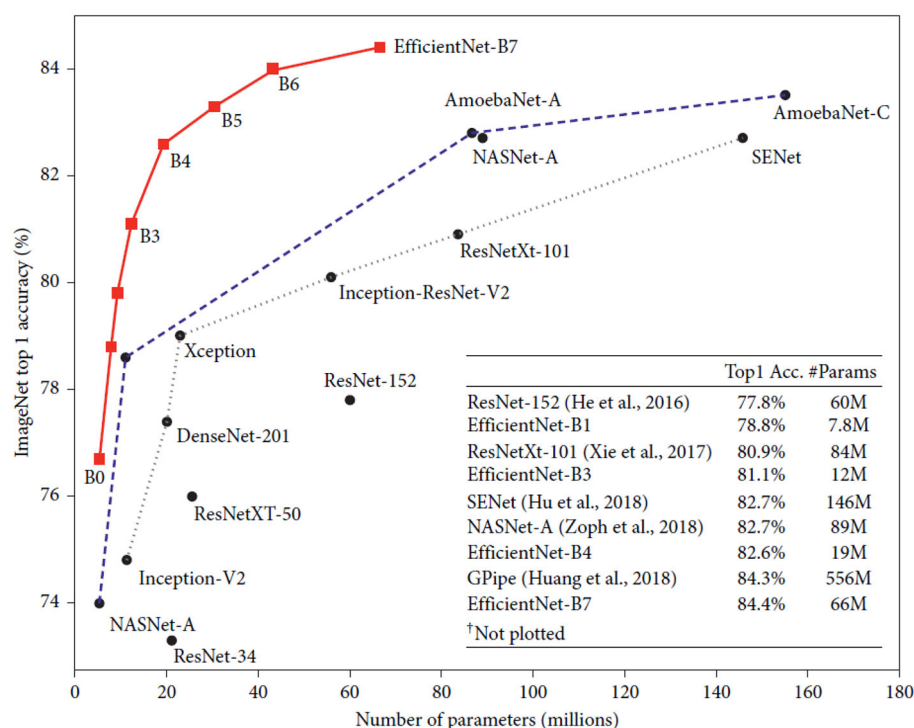


FIGURE 2
Model size vs. ImageNet accuracy (Tan and Le, 2019).

2.7. Methodology

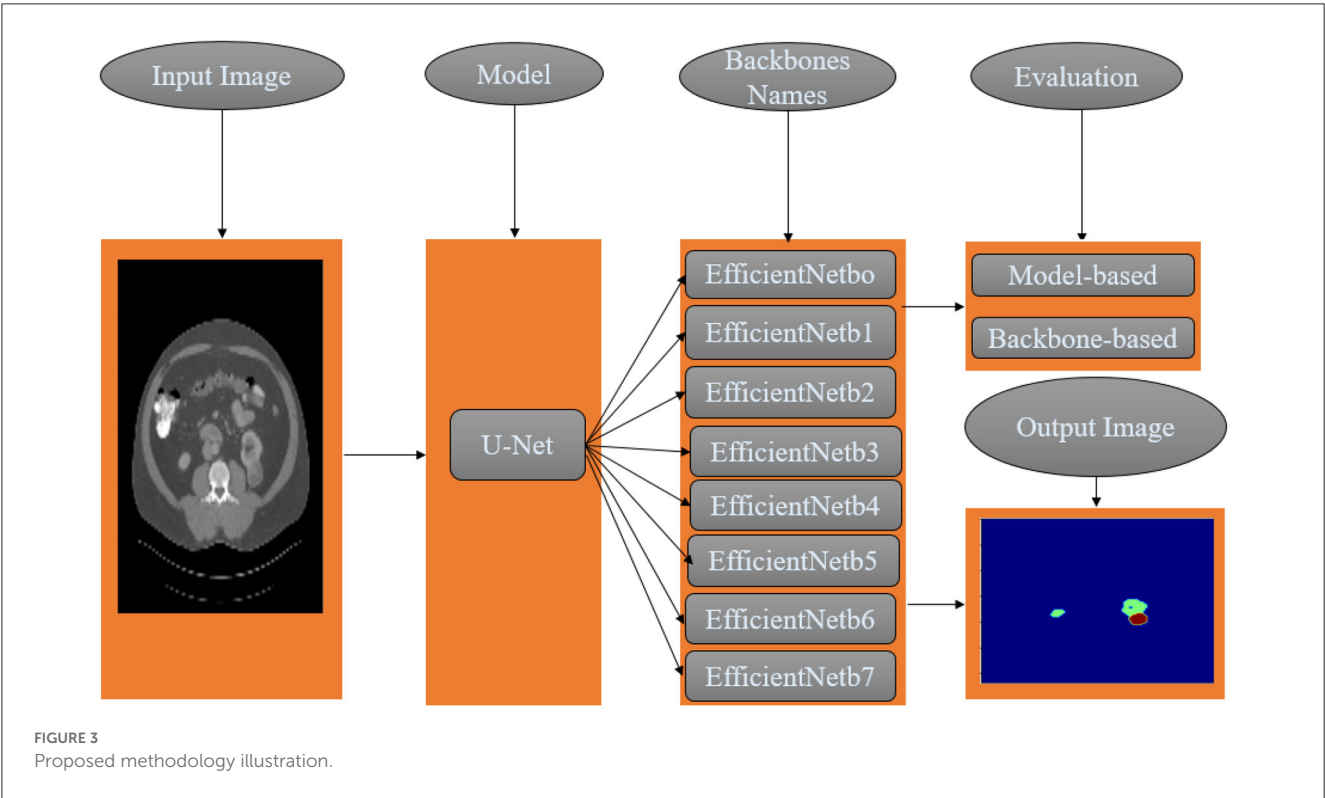
This section explains the suggested method for segmenting tumors and kidneys on CT slices. Slices of CT were obtained from the KiTS19 database to conduct the studies (Heller et al., 2019). Each volume received data preparation in the first stage, which comprised scaling, normalization, and transferring CT volumes from DICOM to png format. Using the EfficientNet U-Net model in the segmentation stage, the kidneys and tumor are segmented concurrently in the second step. The second step of the pipeline involves using an EfficientNet U-Net model to simultaneously segment the kidneys and tumors. The U-Net architecture is commonly used in medical image segmentation tasks because it effectively captures local and global image features. The EfficientNet variant of the U-Net model is designed to be computationally efficient, which makes it suitable for processing large volumes of medical imaging data. Finally, the segmentation results for the kidneys and tumors are obtained from the ensemble phase, where multiple predictions from the EfficientNet U-Net model are combined to produce the best possible segmentation. This is typically done by computing the intersection-over-union (IoU) scores for the different predictions and selecting the segmentation with the highest IoU score.

The IoU_Score measures the overlap between the predicted and ground truth segmentation and is commonly used to evaluate segmentation performance in medical imaging. Segmentation models: this is an open-source library of deep-learning models

for image segmentation tasks. The library implements state-of-the-art architectures for semantic, instance, and panoptic segmentation. Some of the models available in the library include UNet, PSPNet, FPN, LinkNet, and MaskRCNN, among others. The library is built on the Keras deep learning framework and supports TensorFlow and PyTorch backends. It also includes pre-trained weights for various datasets such as Cityscapes, Pascal VOC, and COCO, which can be used for transfer learning or fine-tuning on custom datasets. The segmentation models library is useful for researchers and practitioners working on image segmentation tasks. This can be found on GitHub at https://github.com/qubvel/segmentation_models.

2.8. Backbone architectures

The backbones of feature extraction networks compute image input features, and selecting the optimal network is crucial for objective task performance and Deep Learning (DL) model computational complexity. Numerous backbone networks have been designed and implemented in various DL models. Further research is needed to compare feature extraction networks for DL applications (Elharrouss et al., 2022). This study evaluates eight existing feature extraction backbone networks for a single U-Net model to determine the most effective combination. Unsuitable backbones can degrade performance, be computationally expensive, and be complex.



2.9. Proposed methodology

This study applies semantic segmentation to renal tumors. The proposed methodology combines one semantic segmentation U-Net model and eight feature extraction networks. This modular architecture seeks to identify the optimal segmentation solution for kidney tumors. In addition, the optimal Loss function is examined and evaluated from both the model and backbone perspectives, and a methodology is proposed as shown in Figure 3.

2.10. Implementation

All algorithms in this study were implemented in Python 3.9.16 using Anaconda and Jupyter on a DELL CORE i5 personal computer. Original images (512×512) were resized uniformly to 128×128 pixels for segmentation model input. This section discusses assembling the eight models EfficientNet family “EfficientNetB0”, “EfficientNetB1”, “EfficientNetB2”, “EfficientNetB3”, “EfficientNetB4”, “EfficientNetB5”, “EfficientNetB6”, and “EfficientNetB7”, using U-Net architecture.

2.11. Model setup

Each model is trained for a total of 50 epochs. To avoid overfitting, training ceases when the validation loss remains at or above 0.0001. All backbones have ImageNet-trained weights for accelerated convergence. Table 4 details all DL models’ baseline model configurations and hyperparameters.

TABLE 4 Models’ hyperparameter setup.

Hyperparameter	Settings
Activation	“softmax”
Optimizer	Adam
Learning rate	0.0001
Batch size	32
Epochs	50
Metrics	IoU
Input images size	128×128

3. Results and discussion

This section discusses assembling the eight models EfficientNetB0, EfficientNetB1, EfficientNetB2, EfficientNetB3, EfficientNetB4, EfficientNetB5, EfficientNetB6, and EfficientNetB7 using U-Net architecture. To train all models using EfficientNet U-Net architecture.

3.1. Experimental results

The accuracy of each model using the stain normalization techniques is shown in Table 5, along with the average precision and recall (over the three classes) for each model. Additionally, the EfficientNet model’s average accuracy is determined. In this activity, accuracy and sensitivity (precision) is crucial. As

TABLE 5 Outcomes of the ensemble model of EfficientNetB7.

IoU_Scores	Obtained values
Background IoU_Score	0.999
Kidney IoU_Score	0.977
Tumor IoU_Score	0.962
Mean IoU_Score	0.980

a result, these indicators will be examined and explored in more detail.

3.2. Discussion

3.2.1. Analysis of results

Outcomes revealed that the best-performing models were pertained on ImageNet and fine-tuned. Even though many approaches incorporate patch-wise classification to utilize local features (Aresta et al., 2019), using the entire image for classification produced better results. This insight implies that extracting nuclei and tissue organization features is more valuable for deciphering the image classes than nuclei-scale features. The image-level classification in this research observes the ability of the architecture to extract global features in kidney cancer CT images and use it to classify unseen images Table 6 shows the result.

Convolutional neural network (CNN) models belonging to the EfficientNet family are intended to be efficient in terms of model size and computing cost while maintaining excellent accuracy. The models are intended to be scalable in complexity and size and are trained on big image datasets. Several models in the EfficientNet family are identified by the letters B0–B7, with B0 being the smallest and B7 being the biggest and most sophisticated. The outcomes shown in Table 6 demonstrate how effectively the EfficientNet designs handle the particular image segmentation tasks assessed in this study. The models obtained high IoU_Scores for background, kidney, and tumor segmentation, with mean IoU_Scores ranging from 0.974 for B1 to 0.980 for B4. The chart demonstrates that performance on the segmentation job typically improves when the EfficientNet architecture version grows from B0 to B7. The Mean IoU_Score, for instance, rises from 0.976 for B0 to 0.980 for B4, then slightly falls for B5, B6, and B7. The findings demonstrate that the background segmentation IoU_Score was consistently high, with a score of 0.999, across all EfficientNet designs (B0–B7). B7 received the greatest IoU_Score for segmenting kidneys, scoring 0.977, whereas B4 received the highest IoU_Score for segmenting tumors, scoring 0.968. With mean IoU_Scores ranging from 0.976 for B0 to 0.980 for B4, the results indicate that the EfficientNet designs perform well on these segmentation tasks. These results show that the models can accurately distinguish and segment the required areas of interest in the photos. According to the Table 6, the EfficientNet architecture successfully segments images overall, with higher versions often offering greater performance.

The results presented in Table 6 highlight the exceptional performance of the EfficientNet architectures in handling the specific image segmentation tasks evaluated in this study. The

models achieved impressive Dice similarity coefficient (DSC) Scores for background, kidney, and tumor segmentation, with mean Dice Scores ranging from 0.987 for B1 to 0.990 for B4. Notably, the performance generally improved as the EfficientNet architecture version progressed from B0 to B7. For instance, the Dice Score increased from 0.987 for B1 to 0.990 for B4, then slightly declined for B5, B6, and B7. Furthermore, the background segmentation Dice Scores were consistently high, with a score of 0.999, across all EfficientNet designs (B0–B7). Among them, B7 achieved the highest Dice Score of 0.988 for segmenting kidneys, while B4 obtained the top Dice Score of 0.984 for segmenting tumors. These findings indicate that the EfficientNet designs perform exceptionally well in addressing the challenges of these segmentation tasks.

3.2.2. Computational cost data between the different versions EfficientNetB4 and EfficientNetB7

EfficientNetB4 has a smaller model size compared to EfficientNetB7, which has more layers and parameters, making it larger and more complex. The increased model size allows for more complex patterns and features in data. However, EfficientNetB7 takes longer to train due to its increased complexity and number of parameters. The specific computational cost and performance of these models can vary depending on the implementation, hardware setup, and dataset nature. Table 7 shows the number of parameters and time cost of our model.

According to the Table 8, the (Sharma et al., 2020) model achieved an IoU_Score of 0.97 for kidney and 0.32 for tumor segmentation. On the other hand, “Our Model B4” achieved higher scores of 0.974 for kidney segmentation and 0.968 for tumor segmentation. This suggests that “Our Model B4” outperformed the “Reference” model in accurately segmenting the kidney and tumor regions.

3.2.3. Model accuracy and model loss

The EfficientNet model's performance is evaluated in terms of loss and IoU_Score, indicating its ability to minimize predicted and actual values. The model also measures the overlap between predicted and ground truth bounding boxes. Training and validation IoU_Score and training and validation loss are also measured. For example, the model avoids over-fitting and under-fitting to outlier points in Figures 4, 5.

Table 9 shows the performance of different segmentation approaches on the KiTS19 dataset, where each approach is evaluated based on its kidney and tumor scores. The scores range from 0 to 1, with higher scores indicating better segmentation performance. The approaches listed in the table include various types of convolutional neural networks (CNNs), such as U-Net, V-Net, and RAU-Net, as well as cascaded and boundary-aware networks. Additionally, the table includes the scores for a 3D U-Net and a 3D SEAU-Net. The last two rows of the table represent the scores for “Our approach,” which uses U-Net with two different versions of EfficientNet as its backbone. The first version uses EfficientNet-B4, while the second version

TABLE 6 Results for the EfficientNet architectures using mean IoU, and mean dice.

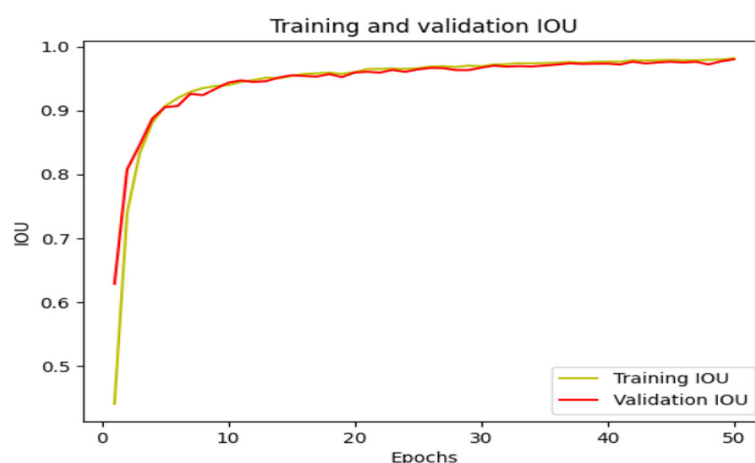
EfficientNet	Background IoU_Score	Kidney IoU_Score	Tumor IoU_Score	Mean IoU_Score	Background dice	Kidney dice	Tumor dice	Mean dice
B0	0.999	0.969	0.960	0.976	0.999	0.984	0.980	0.988
B1	0.999	0.963	0.961	0.974	0.999	0.981	0.980	0.987
B2	0.999	0.966	0.965	0.977	0.999	0.983	0.982	0.988
B3	0.999	0.970	0.960	0.976	0.999	0.985	0.980	0.988
B4	0.999	0.974	0.968	0.980	0.999	0.987	0.984	0.990
B5	0.999	0.972	0.966	0.979	0.999	0.986	0.983	0.989
B6	0.999	0.971	0.963	0.978	0.999	0.985	0.981	0.988
B7	0.999	0.977	0.962	0.980	0.999	0.988	0.981	0.989

TABLE 7 Computational cost data between the different versions EfficientNetB4 and EfficientNetB7.

Versions EfficientNet	Total params	Trainable params	Non-trainable params	Time per second
EfficientNetb4	25.735.307	25.608.123	127.184	114,386
EfficientNetb7	75.048.387	74.735.682	312.704	118,455

TABLE 8 Comparison of object detection results using intersection over union (IoU) with previous studies.

Reference	Dataset	Architecture	Type of images	Kidney IoU_Score	Tumor IoU_Score
Sharma et al. (2020)	KiTS19	U-Net	CT	0.97	0.32
Our Model B4	KiTS19	EfficientNe U-Net	CT	0.974	0.968

FIGURE 4
Model accuracy for EfficientNetB4.

uses EfficientNet-B7. The IoU_scores for Our approach are quite high, with a kidney score of 0.974 and 0.977 and a tumor score of 0.968 and 0.962, respectively. These scores suggest that our approach is highly effective for segmenting kidney and tumor regions in the KiTS19 dataset. Figure 6 compares different segmentation approaches.

4. Conclusion

In conclusion, this paper proposes a method for segmenting tumors and kidneys on CT slices using an EfficientNet U-Net model. The suggested process involves data preparation, segmentation, and an ensemble phase. Preprocessing, down

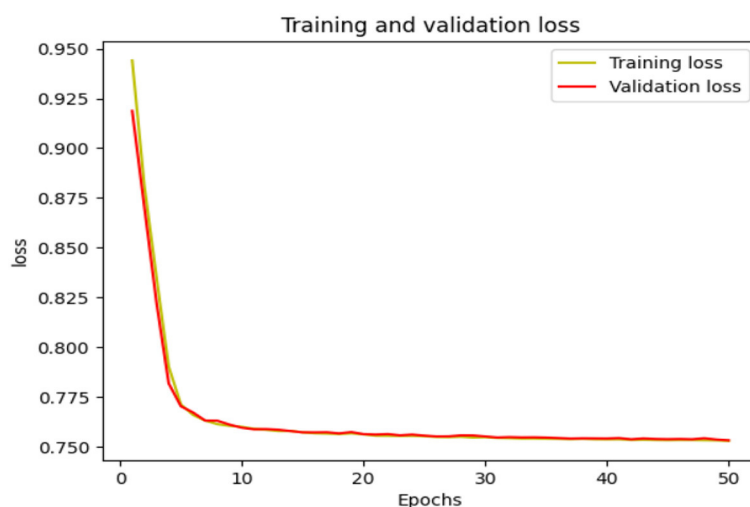


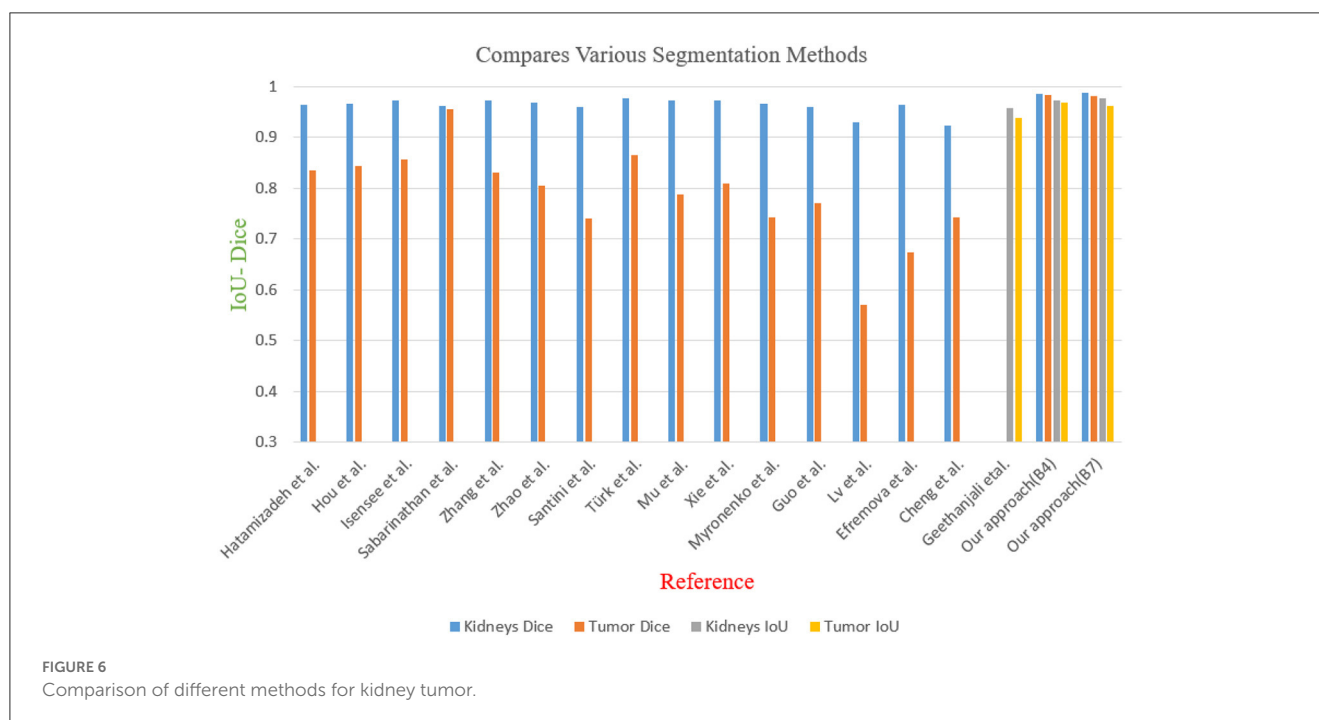
FIGURE 5
Model loss for EfficientNetB4.

TABLE 9 KiTS19 dataset segmentation: evaluating pre-trained architecture approaches against prior methods.

Reference	Architecture	Kidneys dice	Tumor dice	Kidneys IoU scores	Tumor IoU scores
Hatamizadeh et al. (2020)	CNN	0.965	0.835	-	-
Hou et al. (2020)	U-Net	0.967	0.845	-	-
Isensee and Maier-Hein (2019)	3D U-Net	0.974	0.856	-	-
Sabarinathan et al. (2020)	U-Net	0.963	0.955	-	-
Zhang et al. (2019)	Cascaded	0.974	0.831	-	-
Zhao et al. (2020)	U-Net	0.969	0.805	-	-
Santini et al. (2019)	Multi-stage U-Net	0.96	0.74	-	-
Türk et al. (2020)	V- Net	0.977	0.865	-	-
Mu et al. (2019)	V- Net	0.974	0.789	-	-
Myronenko and Hatamizadeh (2019)	Boundary-Aware	0.974	0.810	-	-
Xie et al. (2020)	Cascaded	0.967	0.743	-	-
Guo et al. (2021)	RAU-Net	0.960	0.770	-	-
Lv and Wang (2021)	3D U-Net	0.930	0.570	-	-
Efremova et al. (2019)	CNN	0.964	0.674	-	-
Cheng et al. (2019)	3D SEAU -Net	0.924	0.743	-	-
Geethanjali and Dinesh (2021)	U-Net	-	-	0.957	0.939
Our approach	U-net EfficientNet-B4	0.987	0.984	0.974	0.968
Our approach	U-net-EfficientNet-B7	0.988	0.981	0.977	0.962

sampling, and resizing CT images are also essential to classifying CT images. The paper also discusses the IoU metric as a common evaluation metric for image segmentation tasks. Finally, the paper summarizes related techniques in the literature and the U-Net architecture used for image segmentation tasks. The proposed method can improve the segmentation accuracy and the diagnosis of tumors and kidneys on CT slices.

This study explored the performance of the U-Net architecture when combined with various models of the EfficientNet family for image segmentation in kidney cancer CT images. The results showed that the models fine-tuned on ImageNet outperformed patch-wise classification approaches regarding accuracy and sensitivity. The EfficientNet models showed excellent background, kidney, and tumor segmentation results, with mean IoU_Scores



ranging from 0.976 to 0.980. In particular, B4 achieved the highest IoU_Score for segmenting tumors, with a score of 0.968, while B7 received the highest IoU_Score for segmenting kidneys, with a score of 0.977. Comparison with previous studies revealed that the models presented in this study achieved higher IoU_Scores for kidney and tumor segmentation. These results suggest that the U-Net architecture combined with EfficientNet models can accurately distinguish and segment the required areas of interest in kidney cancer CT images.

The last table presents the performance of different segmentation approaches for kidney and tumor regions in the KiTS19 dataset, where higher scores indicate better segmentation performance. The table includes several types of convolutional neural networks, such as U-Net, V-Net, and RAU-Net, alongside cascaded and boundary-aware networks. "Our approach," which utilizes U-Net with two different versions of EfficientNet as its backbone, achieves impressive Dice scores of 0.987 and 0.988 for the kidney and 0.984 and 0.981 for the tumor scores. These results suggest that our approach is a highly effective segmentation technique for the KiTS19 dataset.

Based on the findings and conclusions of this study, there are several potential future directions for research in this area. Some possible areas for future work include further optimization of the proposed method: the proposed method in this paper achieved high accuracy in segmenting tumors and kidneys on CT slices. However, there is still room for further optimization and improvement of the method. For example, the authors could explore using different pre-processing techniques, different data augmentation strategies, or other modifications to the U-Net architecture to improve performance further. Application to other medical imaging datasets: the proposed method was evaluated on the KiTS19 dataset, a benchmark dataset for kidney tumor segmentation. However, it would be interesting to see how well

the method performs on other medical imaging datasets, such as lung, brain, or liver imaging. Comparing the performance of the proposed method with other state-of-the-art techniques on different datasets could help to establish its generalizability and robustness. Exploration of other deep learning architectures: the authors of this paper used the U-Net architecture combined with EfficientNet models for image segmentation tasks. However, many other deep learning architectures, such as Attention U-Net, DeepLab, or Mask R-CNN, could be explored for this purpose. Future work could explore using different architectures and compare their performance with the proposed method.

Data availability statement

Publicly available datasets were analyzed in this study. The original dataset is in <https://kits19.grand-challenge.org/data/>, and we used the dataset found at <https://www.kaggle.com/competitions/body-morphometry-kidney-and-tumor/data>.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

References

- Abdelrahman, A., and Viriri, S. (2022). Kidney tumor semantic segmentation using deep learning: a survey of state-of-the-art. *J. Imag.* 8, 55. doi: 10.3390/jimaging8030055
- Ali, A. M., Farag, A. A., and El-Baz, A. S. (2007). "Graph cuts framework for kidney segmentation with prior shape constraints," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2007: 10th International Conference, Brisbane, Australia, October 29–November 2, 2007, Proceedings, Part I 10* (Berlin Heidelberg: Springer) 384–392. doi: 10.1007/978-3-540-75757-3_47
- Aresta, G., Araújo, T., Kwok, S., Chennamsetty, S. S., Safwan, M., Alex, V., et al. (2019). BACH: Grand challenge on breast cancer histology images. *Med. Image Anal.* 56, 122–139. doi: 10.1016/j.media.2019.05.010
- Avilov, O., Rimbart, S., Popov, A., and Bougrain, L. (2020). "Deep learning techniques to improve intraoperative awareness detection from electroencephalographic signals," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 142–145*. doi: 10.1109/EMBC44109.2020.9176228
- Bapir, R., Hammood, Z. D., Omar, S. S., Salih, A. M., Kakamad, F. H., Najjar, K. A., et al. (2022). Synchronous invasive ductal breast cancer with clear cell renal carcinoma: a rare case report with review of literature. *IJS Short Rep.* 7, e59. doi: 10.1097/SR9.0000000000000059
- Beisland, C., Talleraas, O., Bakke, A., and Norstein, J. (2006). Multiple primary malignancies in patients with renal cell carcinoma: A national population-based cohort study. *BJU Int.* 97, 698–702. doi: 10.1111/j.1464-410X.2006.06004.x
- Causey, J., Stubblefield, J., Qualls, J., Fowler, J., Cai, L., Walker, K., et al. (2021). An ensemble of U-Net models for kidney tumor segmentation with CT images. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 5963, 1–5. doi: 10.1109/TCBB.2021.3085608
- Chen, Z., and Liu, H. (2021). "5 D cascaded semantic segmentation for kidney tumor Cyst," in *International Challenge on Kidney and Kidney Tumor Segmentation* (Cham: Springer International Publishing) 28–34. doi: 10.1007/978-3-030-98385-7_4
- Cheng, J., Liu, J., Liu, L., Pan, Y., and Wong, J. (2019). *A Double Cascaded Framework Based on 3D SEAU-Net for Kidney and Kidney Tumor Segmentation*. University of Minnesota Libraries Publishing. doi: 10.24926/548719.067
- Ciregan, D., Meier, U., and Schmidhuber, J. (2012). "Multi-column deep neural networks for image classification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 3642–3649. doi: 10.1109/CVPR.2012.6248110
- Collobert, R., and Weston, J. (2008). "A unified architecture for natural language processing: Deep neural networks with multitask learning," *Proceedings of the 25th International Conference on Machine Learning* 160–167. doi: 10.1145/1390156.1390177
- da Cruz, L. B., Araújo, J. D. L., Ferreira, J. L., Diniz, J. O. B., Silva, A. C., de Almeida, J. D. S., et al. (2020). Kidney segmentation from computed tomography images using deep neural network. *Comput. Biol. Med.* 123, 103906. doi: 10.1016/j.combiomed.2020.103906
- Efremova, D. B., Kononov, D. A., Siriapisith, T., Kusakunniran, W., and Haddawy, P. (2019). Automatic segmentation of kidney and liver tumors in CT images. *arXiv preprint arXiv:1908.01279*. doi: 10.24926/548719.038
- Elharrouss, O., Akbari, Y., Almaadeed, N., and Al-Maadeed, S. (2022). Backbones-review: Feature extraction networks for deep learning and deep reinforcement learning approaches. *arXiv preprint arXiv:2206.08016*.
- Geethanjali, T. M., and Dinesh, M. S. (2021). "Semantic segmentation of tumors in kidneys using attention U-Net models," in *2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques, ICEECCOT 2021 - Proceedings* 286–290. doi: 10.1109/ICEECCOT52851.2021.9708025
- Guo, J., Zeng, W., Yu, S., and Xiao, J. (2021). "RAU-Net: U-Net model based on residual and attention for kidney and kidney tumor segmentation," in *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)* (IEEE) 353–356. doi: 10.1109/ICCECE51280.2021.9342530
- Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., and Yener, B. (2009). Histopathological image analysis: a review. *IEEE Rev. Biomed. Eng.* 2, 147–171. doi: 10.1109/RBME.2009.2034865
- Hatamizadeh, A., Terzopoulos, D., and Myronenko, A. (2020). Edge-Gated CNNs for volumetric semantic segmentation of medical images. *arXiv preprint arXiv:2002.04207*.
- Hatipoglu, N., and Bilgin, G. (2017). Cell segmentation in histopathological images with deep learning algorithms by utilizing spatial relationships. *Med. Biol. Eng. Comput.* 55, 1829–1848. doi: 10.1007/s11517-017-1630-1
- Heller, N., Sathianathan, N., Kalapara, A., Walczak, E., Moore, K., Kaluzniak, H., et al. (2019). The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. *arXiv preprint arXiv:1904.00445*.
- Hou, X., Xie, C., Li, F., Wang, J., Lv, C., Xie, G., et al. (2020). "A triple-stage self-guided network for kidney tumor segmentation," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* (IEEE) 341–344. doi: 10.1109/ISBI45749.2020.9098609
- Hsiao, C.-H., Lin, P.-C., Chung, L.-A., Lin, F. Y.-S., Yang, F.-J., Yang, S.-Y., et al. (2022). A deep learning-based precision and automatic kidney segmentation system using efficient feature pyramid networks in computed tomography images. *Comput. Methods Progr. Biomed.* 221, 106854. doi: 10.1016/j.cmpb.2022.106854
- Isensee, F., and Maier-Hein, K. H. (2019). An attempt at beating the 3D U-Net. *arXiv preprint arXiv:1908.02182*.
- Kandel, I., and Castelli, M. (2020). How deeply to fine-tune a convolutional neural network: a case study using a histopathology dataset. *Appl. Sci.* 10, 3359. doi: 10.3390/app10103359
- Khan, A., Sohail, A., Zahoor, U., and Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artif. Intell. Rev.* 53, 5455–5516. doi: 10.1007/s10462-020-09825-6
- Kim, K. G. (2019). Deep learning book review. *Nature* 29, 1–73. doi: 10.4258/hir.2016.22.4.351
- Lv, Y., and Wang, J. (2021). "Three uses of one neural network: automatic segmentation of kidney tumor and cysts based on 3D U-Net," in *International Challenge on Kidney and Kidney Tumor Segmentation* (Cham: Springer International Publishing) 40–45. doi: 10.1007/978-3-030-98385-7_6
- Mayer, K. H., Safren, S. A., Elsesser, S. A., Psaros, C., Tinsley, J. P., Marzinke, M., et al. (2017). Optimizing pre-exposure antiretroviral prophylaxis adherence in men who have sex with men: results of a pilot randomized controlled trial of "Life-Steps for PrEP". *AIDS Behav.* 21, 1350–1360. doi: 10.1007/s10461-016-1606-4
- Mu, G., Lin, Z., Han, M., Yao, G., and Gao, Y. (2019). *Segmentation of Kidney Tumor by Multi-Resolution VB-Nets*. University of Minnesota Libraries Publishing. doi: 10.24926/548719.003
- Munien, C., and Viriri, S. (2021). Classification of hematoxylin and eosin-stained breast cancer histology microscopy images using transfer learning with EfficientNets. *Comput. Intell. Neurosci.* 2021, 5580914. doi: 10.1155/2021/5580914
- Myronenko, A., and Hatamizadeh, A. (2019). 3D kidneys and kidney tumor semantic segmentation using boundary-aware networks. *arXiv preprint arXiv:1909.06684*.
- Nazari, M., Jiménez-Franco, L. D., Schroeder, M., Kluge, A., Bronzel, M., and Kimiaei, S. (2021). Automated and robust organ segmentation for 3D-based internal dose calculation. *EJNMMI Res.* 11, 1–13. doi: 10.1186/s13550-021-00796-5
- Pandey, S., Singh, P. R., and Tian, J. (2020). Biomedical Signal Processing and Control An image augmentation approach using two-stage generative adversarial network for nuclei image segmentation. *Biomed. Signal Proc. Control* 57, 101782. doi: 10.1016/j.bspc.2019.101782
- Payer, C., Štern, D., Bischof, H., and Urschler, M. (2016). "Regressing heatmaps for multiple landmark localization using CNNs," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Cham: Springer International Publishing) 230–238. doi: 10.1007/978-3-319-46723-8_27
- Rajinikanth, V., Vincent, P. M. D. R., Srinivasan, K., Prabhu, G. A., and Chang, C.-Y. (2023). A framework to distinguish healthy/cancer renal CT images using the fused deep features. *Front. Public Health* 11, 1109236. doi: 10.3389/fpubh.2023.1109236
- Ramadan, S. Z. (2020). Methods used in computer-aided diagnosis for breast cancer detection using mammograms: a review. *J. Healthcare Eng.* 2020, 9162464. doi: 10.1155/2020/9162464
- Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany,*

October 5-9, 2015, *Proceedings, Part III* 18 (Cham: Springer International Publishing) 234–241. doi: 10.1007/978-3-319-24574-4_28

Sabarinathan, D., Parisa Beham, M., and Mansoor Roomi, S. M. M. (2020). Hyper vision net: kidney tumor segmentation using coordinate convolutional layer and attention unit. *Commun. Comput. Inform. Sci.* 1249, 609–618. doi: 10.1007/978-981-15-8697-2_57

Santini, G., Moreau, N., and 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*. doi: 10.24926/548719.023

Sharma, R., Halarankar, P., and Choudhari, K. (2020). “Kidney and tumor segmentation using U-Net deep learning model,” in *5th International Conference on Next Generation Computing Technologies (NGCT-2019)*. doi: 10.2139/ssrn.3527410

Tan, M., and Le, Q. (2019). “Efficientnet: Rethinking model scaling for convolutional neural networks,” in *International Conference on Machine Learning (PMLR)* 6105–6114.

Tan, M., and Le, Q. V. (2021). “EfficientNetV2: smaller models and faster training,” in *International Conference on Machine Learning (PMLR)* 10096–10106.

Thong, W., Kadoury, S., Piché, N., and Pal, C. J. (2018). Convolutional networks for kidney segmentation in contrast-enhanced CT scans. *Comput. Methods Biomech. Biomed. Eng.* 6, 277–282. doi: 10.1080/21681163.2016.1148636

Torres, H. R., Queirós, S., Morais, P., Oliveira, B., Fonseca, J. C., and Vilaça, J. L. (2018). Kidney segmentation in ultrasound, magnetic resonance and computed tomography images: A systematic review. *Comput. Methods Programs Biomed.* 157, 49–67. doi: 10.1016/j.cmpb.2018.01.014

Tsantekidis, A., Passalis, N., Tefas, A., Kannianen, J., Gabbouj, M., and Iosifidis, A. (2017). “Forecasting stock prices from the limit order book using convolutional neural networks,” in *2017 IEEE 19th Conference on Business Informatics (CBI)* (IEEE) 7–12. doi: 10.1109/CBI.2017.23

Tsuneki, M. (2022). Deep learning models in medical image analysis. *J. Oral Biosci.* 64, 312–320. doi: 10.1016/j.job.2022.03.003

Türk, F., Lüy, M., and Bar, işçi, N. (2020). Kidney and renal tumor segmentation using a hybrid v-net-based model. *Mathematics* 8, 1–17. doi: 10.3390/math8101772

Üreyen, O., Dadal, i, E., Akdeniz, F, Sahin, T., Tekeli, M. T., Eliyatkin, N., et al. (2015). Co-existent breast and renal cancer. *Turkish J. Surg.* 31, 238–240. doi: 10.5152/UCD.2015.2874

Valueva, M. V., Nagornov, N. N., Lyakhov, P. A., Valuev, G. V., and Chervyakov, N. I. (2020). Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. *Mathem. Comput. Simul.* 177, 232–243. doi: 10.1016/j.matcom.2020.04.031

Wen, J., Li, Z., Shen, Z., Zheng, Y., and Zheng, S. (2021). “Squeeze-and-excitation encoder-decoder network for kidney and kidney tumor segmentation in CT images,” in *International Challenge on Kidney and Kidney Tumor Segmentation* (Cham: Springer International Publishing) 71–79. doi: 10.1007/978-3-030-98385-7_10

Xie, X., Li, L., Lian, S., Chen, S., and Luo, Z. (2020). SERU: A cascaded SE-ResNeXT U-Net for kidney and tumor segmentation. *Concurr. Comput.* 32, e5738. doi: 10.1002/cpe.5738

Yang, G., Li, G., Pan, T., Kong, Y., Wu, J., Shu, H., et al. (2018). “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)* (IEEE) 3790–3795. doi: 10.1109/ICPR.2018.8545143

Zhang, Y., Wang, Y., Hou, F., Yang, J., Xiong, G., Tian, J., et al. (2019). Cascaded volumetric convolutional network for kidney tumor segmentation from CT volumes. *arXiv preprint arXiv:1910.02235*. doi: 10.24926/548719.004

Zhao, W., Jiang, D., Queralta, J. P., and Westerlund, T. (2020). MSS U-Net: 3D segmentation of kidneys and tumors from CT images with a multi-scale supervised U-Net. *Inform. Med. Unlocked* 19, 100357. doi: 10.1016/j.imu.2020.100357

3.2 Conclusion

This paper "EfficientNet Family U-Net Models for Deep Learning Semantic Segmentation of Kidney Tumors on CT Images" demonstrates the effectiveness of combining the EfficientNet architecture with U-Net for precise and efficient kidney tumor segmentation. The EfficientNet family, known for its balance between accuracy and computational efficiency, enhances the segmentation capability by optimizing model size and performance. The integration with U-Net allows for improved handling of spatial information and fine details in CT images, crucial for accurate tumor identification. The study shows that the EfficientNet U-Net models outperform conventional architectures in terms of accuracy, while also being resource-efficient, making them particularly suitable for clinical applications where real-time and reliable tumor segmentation is essential. The paper also suggests that these models can potentially be adapted for other medical imaging tasks, further broadening their impact in the field of medical diagnostics.

Chapter 4

FPN-SE-ResNet Model for Accurate Diagnosis of Kidney Tumors Using CT Images

4.1 Introduction

The paper "FPN-SE-ResNet Model for Accurate Diagnosis of Kidney Tumors Using CT Images" introduces a novel deep learning model that combines Feature Pyramid Network (FPN), Squeeze-and-Excitation (SE) blocks, and ResNet architecture to enhance the accuracy of kidney tumor diagnosis from CT scans. By integrating these components, the model effectively captures multi-scale features while emphasizing important image regions, leading to improved performance in tumor detection and segmentation. The study demonstrates the model's ability to deliver more precise diagnoses compared to traditional approaches, highlighting its potential for clinical application in kidney tumor analysis.

Article

FPN-SE-ResNet Model for Accurate Diagnosis of Kidney Tumors Using CT Images

Abubaker Abdelrahman  and Serestina Viriri * 

School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal,
Durban 4000, South Africa; 220106722@stu.ukzn.ac.za

* Correspondence: viriris@ukzn.ac.za

Abstract: Kidney tumors are a significant health concern. Early detection and accurate segmentation of kidney tumors are crucial for timely and effective treatment, which can improve patient outcomes. Deep learning techniques, particularly Convolutional Neural Networks (CNNs), have shown great promise in medical image analysis, including identifying and segmenting kidney tumors. Computed tomography (CT) scans of kidneys aid in tumor assessment and morphology studies, employing semantic segmentation techniques for precise pixel-level identification of kidneys and surrounding anatomical structures. This paper proposes a Squeeze-and-Excitation-ResNet (SE-ResNet) model for segmentation by combining the encoder stage of SE-ResNet with the Feature Pyramid Network (FPN). The performance of the proposed SE-ResNet model is evaluated using the Intersection over Union (IoU) and F1-score metrics. Experimental results demonstrate that the SE-ResNet models achieve impressive IoU scores for background, kidney, and tumor segmentation, with mean IoU scores ranging from 0.988 to 0.981 for Seresnet50 and Seresnet18, respectively. Notably, Seresnet50 exhibits the highest IoU score for kidney segmentation. These findings suggest that SE-ResNet models accurately identify and segment regions of interest in CT images of renal carcinoma, with higher model versions generally exhibiting superior performance. The proposed Seresnet50 model is a good tool for accurate tumor detection and image classification, aiding medical professionals in early diagnosis and timely intervention.



Citation: Abdelrahman, A.; Viriri, S. FPN-SE-ResNet Model for Accurate Diagnosis of Kidney Tumors Using CT Images. *Appl. Sci.* **2023**, *13*, 9802. <https://doi.org/10.3390/app13179802>

Academic Editor: Yudong Zhang

Received: 4 August 2023

Revised: 25 August 2023

Accepted: 29 August 2023

Published: 30 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: Feature Pyramid Network; SE-ResNet; kidney tumors; diagnosis; semantic segmentation

1. Introduction

1.1. Diagnosis

Diagnosis is the medical process of identifying the cause and effect of a disease or disorder using logic, analytics, and experience. It involves analyzing a patient's medical history and physical examination to inform diagnosis, influencing the doctor's visit process. Diagnosis involves statistical classification evaluations and categorization of an individual's condition into distinct categories for medical decisions regarding treatment and prognosis. A diagnostic opinion is often described as a disease or other condition. Diagnosis is a crucial aspect of clinical decision-making, often involving ambiguity and uncertainty due to the inherent nature of the medical field [1–4]. The disease consists of observable signs and symptoms that medical professionals must interpret. The diagnosis procedure involves interpretation, a complex decision-making process in general medicine to accurately understand a patient's health problem [5]. Disease diagnosis is crucial for clinical decision-making, involving subjective and objective factors. An accurate and prompt diagnosis is essential in determining a disease or disorder. A definitive diagnosis must be established before a treatment plan can be developed [6]. Diagnosis is a complex and challenging process for healthcare professionals, requiring them to consider multiple factors and circumstances. They gather empirical data to manage patients' problems and reduce diagnostic uncertainty, ultimately improving healthcare quality [6]. Disease diagnosis is

ambiguous and complex due to patient susceptibility and the complexity of the medical process. Unintended errors can occur due to the ambiguous nature of the disease, patient information, and the inherent essence of medicine [7,8].

1.2. Computer-Aided Diagnosis

Numerous researchers have utilized computer-assisted methods to enhance disease diagnosis to aid physicians in making the most accurate decisions, as accurate diagnosis is a crucial aspect of healthcare subjects [9–14]. Computer-aided diagnosis (CAD) can be defined as a radiologist's use of output from a computer analysis of medical images as a "second opinion" in detecting lesions and making diagnostic decisions. The radiologist can then make the definitive diagnosis and treatment decisions. Recently, At many screening sites and hospitals in the United States, CAD has become an integral element of the routine clinical work for detecting cancer on medical images. CAD is increasingly being utilized to detect and diagnose abnormalities in medical images from various imaging modalities, making it a crucial research topic in medical imaging and diagnostic radiology [15–19].

1.3. Medical Imaging

Medical imaging techniques like X-rays, MRI, and CT require subjective analysis by radiologists. With increasing techniques, manual analysis becomes time-consuming. Machine learning models mimic human visual perception, enabling the automatic classification of images as diagnostic aids. Advances in AI and computing hardware have made computer-assisted methods more powerful, making them useful for diagnosticians. Consequently, CAD has acquired theoretical and practical significance as a significant trend in medical science. Using computer vision to automatically analyze and process medical images has many distinct benefits [13,20–22]. Utilizing medical images like PET, SPECT, MRI, and CT offers vital insights into head and neck cancer's dimensions, site, structure, and metabolism [23]. Navigating clinical outcome prediction proves intricate yet pivotal for guiding treatment strategies [24]. In contrast to standard regression tasks like predicting clinical scores, clinical outcome prediction presents greater complexity due to survival data censoring, causing event times to be indeterminate for certain patients [25]. Accurate kidney tumor segmentation is pivotal in advancing kidney tumor diagnosis and treatment strategies through early detection, treatment planning, and risk assessment. However, real-time medical image analysis faces challenges like poor image quality, diverse protocols, and patient variations. Modern hardware's computational power enables rapid analysis and processing, preventing fatigue and cognitive issues. Computer technologies facilitate data transmission, enabling accurate diagnosis of patients in remote areas. Machine learning algorithms and artificial intelligence classifiers improve prediction accuracy for kidney stone patients [20]. Since the introduction of the first (CNN) (LeNet-5 [26]) in 1989, CNNs exhibited exceptional performance in medical image processing [27,28]. The evolution of these models has also contributed to the precision of CAD techniques [17,29].

1.4. Convolutional Neural Network

(CNN) has accomplished some outstanding feats. In the area of deep learning, it has grown to be one of the most representative neural networks. People have been able to do things previously thought impossible, like facial recognition, driverless automobiles, self-service supermarkets, and intelligent medical treatments, because of computer vision based on CNN [30]. Recent developments in deep convolutional neural networks (CNNs) have showcased their exceptional capabilities in mastering segmentation tasks across various imaging modalities and for diverse anatomical structures, e.g., prostate [31,32], heart [33,34], brain [35,36], head and neck [37,38], and COVID-19 [39]. The domain of automated medical image segmentation has witnessed remarkable advancements, reaching the pinnacle of performance through the utilization of deep learning and convolutional neural networks (CNNs) [27]. Over the past few years, convolutional neural network (CNN) based models for medical image segmentation have garnered substantial prowess, rivalling

the performance levels achieved by radiologists [27,40]. However, these models have functioned as autonomous applications, tailored with refined architectures, preprocessing methodologies, data augmentation strategies, and metrics meticulously crafted to suit the distinct characteristics of their datasets and corresponding segmentation challenges [41]. Convolutional neural networks (CNNs) have emerged as the predominant deep learning model for organ and lesion segmentation tasks [42].

1.5. Kidney Cancer

The kidneys in the abdominal region are essential for bodily functions, containing connective tissue and fat. They dispose of metabolized compounds, maintain water and salt balance, and produce hormones for red blood cell production and blood pressure regulation. Figure 1 presents a visual depiction of the human kidneys. However, risk factors like smoking, obesity, and hypertension can compromise kidney function, leading to kidney cancer. DNA alterations can induce kidney cells to deviate from their fundamental role. Early detection is crucial for successful treatment and increased survival chances. Medical imaging tools like CT reveal the complex abdominal interior structure, often used for early-stage cancer detection [43,44]. Renal cancer ranks tenth globally, with a significant increase in individuals. In the US, 2023 projections predict 81,800 new RC diagnoses, resulting in 14,890 deaths [45]. Renal cancer is a diverse group of tumors originating from different kidney cell types, with renal cell carcinoma (RCC) being the most aggressive form, accounting for 70% of all RC cases [46]. The prevalent subtypes of renal cell carcinomas (RCCs) encompass clear cell, chromophobe, oncocytoma, papillary, and other RCC subcategories. Accurate classification of histological subtypes within renal cell carcinoma is paramount to prevent unnecessary biopsies or surgeries [47]. By the World Health Organization (WHO) [48], The classification of RCC subtypes is of great importance as each type has its prognosis [48]. Conventional diagnostic methods may lead to misclassification of benign lesions like angiomyolipoma (AML) and oncocytoma (ONC) as RCC [49]. Incorrectly identifying benign lesions can lead to unwarranted surgical interventions, with 15–20% of tumors surgically removed with a preoperative RCC diagnosis potentially being angiomyolipoma (AML) [50]. Early and accurate diagnosis of renal tumors is crucial for treatment. Conventional tests like CBC, urinalysis, and blood assays can hint at RC but do not provide precise diagnoses or differentiate between subtypes, grades, or stages. Therefore, biopsy remains the ultimate benchmark for definitive RC diagnosis, as it accurately quantifies red blood cells, identifies blood, bacteria, or malignant cells, and gauges renal function [45]. Deep learning is a powerful machine learning technology that can autonomously acquire various features and patterns without human intervention [51,52]. DL has revolutionized prognostic models for early tumor detection using pattern analysis techniques, demonstrating superior performance over conventional machine learning due to its exceptional precision in delivering results [53]. Object detection, a technique used in image processing, has gained significant interest in medical radiology due to its focus on extracting valuable insights from images. This concept includes single-class and multiclass object detection, crucial in identifying object categories [54]. DL methodologies have significantly improved performance and automated segmentation models, especially with the latest advancements in convolutional neural networks (CNNs) [52,55]. CNNs are ideal for medical imaging because they can discern nuanced details. The SE-ResNet architecture enhances feature representation, highlighting relevant image structures and enabling CNNs to outperform traditional methods for more precise and efficient segmentation. In this study, the investigation delves into the utilization of five architectures from the SE-ResNet lineage. Additionally, the study involves a comparison of outcomes across different SE-ResNet backbones.

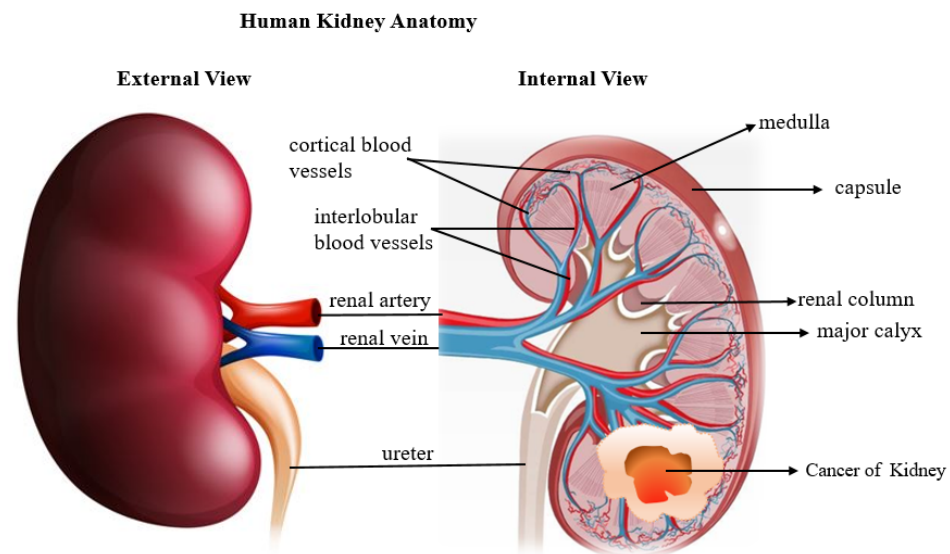


Figure 1. Diagram depicting human kidney anatomy and renal cell carcinoma development [13].

1.6. Current Methods

Numerous DL techniques for kidney segmentation have been created utilizing (CNNs). In reference [56], The proposed 3D fully convolutional network uses a pyramid combining unit and progressively improved feature unit, evaluated using a specific dataset with an average Dice coefficient of 0.931%, by Yang et al. Cruz et al. [57] AlexNet was utilized as a classifier to distinguish non-kidney CT slices from kidney slices to prevent segment discontinuity, and a phase of recovery was implemented. Later, in initial segmentation by a 2D U-Net architecture, the CCL method decreased the number of false-positive predictions. A local test dataset produced a Dice coefficient degree of 96.33% for the (KiTS19) dataset. Hou et al. [58] Instead of aggregating operations, dilated convolution blocks were utilized in a three-stage 3D U-Net. Due to the three phases within the 3D U-Net, input images comprised low-resolution, high-resolution, and trimmed images. A hybrid reduction function comprised of (DC) and a weighted cross-entropy was employed. On the KiTS19 dataset, the technique attained an average Dice score(DS) of 0.9674. Türk et al. [59] A hybrid 3D V-Net model using ResNet++ output layer architecture outperformed the original V-Net with a 0.97 Dice coefficient. Training is challenging due to time and computing resources, and lightweight 2D models are preferred for kidney segmentation assignments.

1.7. Problem Description

Methods of segmentation based on the proximity of pixels are region-based. The authors in reference [60] Propose coarse-to-fine kidney segmentation using kernel fact hunt and adaptive region-growing, modifying attention construction limitations. Erdt et al. [61] Utilizing native restraints prevents adjacent organs from affecting deformable models. Khalifa et al. [62] The level-set procedure generates a deformable model, probabilistic form, voxel communication, and stochastic haste purpose for segmentation presentation; the atlas-based technique uses functional diagrams for orientation and consistent comments. Yang et al. [63] Two-step multi-atlas image registration technique for renal segmentation using random forest and machine learning classifier. Cu-ingnet et al. [64] Random forest algorithm classifies dynamic contrast-enhanced images. Jinet et al. [65,66] Earlier attempts to resolve this issue have shown result promise. This study uses five lightweight architectures from the SE-ResNet family designed to optimize resources while maintaining high precision. These architectures are valuable for achieving comparable outcomes to current methods while requiring less space and training time. The study evaluates SE-ResNets' ability to classify kidney tumor images using F1-score, IoU scores, and the KiTs 19 dataset.

1.8. Paper Structure

This paper is organized into different sections. Section 2, the Materials and Methods, discusses the Dataset and provides information on the framework used in this study. Section 3, Results and Discussion. Finally, Section 4 discusses the insights gained from this study and concludes the paper.

2. Materials and Methods

This paper uses part of the KiTS19 dataset, including CT scans of abdomens with kidney tumors, to train and evaluate segmentation algorithms for detecting and delineating kidney tumours. The study employs Intersection over Union (IoU) and F1-score metrics to assess the method's accuracy and effectiveness. The primary objective is to achieve accurate and efficient identification and delineation of kidney tumors and kidneys within CT images, supporting clinicians in diagnosing and monitoring kidney tumors and improving patient care and treatment outcomes.

2.1. Dataset

The facts for the examination are obtained from KiTS19 [67]. It is a dataset for training and testing machine learning models that comprises CT scans and corresponding annotations for kidney tumor segmentation. It is a widely utilized benchmark dataset for assessing the performance of medical image segmentation models for renal tumors. The testing and challenging training datasets consist of 210 CT volumes of the abdomen. Each volume's imaging and ground truth labels are provided in Neuroimaging Informative Technology Informative (DICOM) format with shape (number of segments, height, and width). It is located on Kaggle at "<https://www.kaggle.com/competitions/body-morphometry-kidney-and-tumor/data/>" (accessed on 1 March 2023) (Who used the original data in reference [67]) The dataset consists of segments ranging in size from 29 to 1059 pixels and grayscale images measuring 512×512 pixels. The characteristics of the experimental dataset are provided in Table 1 below. This study utilized a subset of the dataset of 7899 PNG images.

Table 1. Dataset Properties Used for the Experimentation are cited.

Properties	Values
Number of all Image	7899
Number of Training Images	5841
Number of Validation Images	1027
Number of Testing Images	1031
Image format	Png
Modality	CT

2.2. Preprocessing

Indeed, preprocessing is essential to the classification of CT images. Typically, CT images are stained with various dyes to emphasize various structures and characteristics, and they can vary considerably in terms of staining intensity, colour, and texture. Therefore, preprocessing is required to improve image quality and consistency and eliminate noise and artefacts that could interfere with the classification process [68]. In contrast, Convolutional neural networks (CNNs) are designed to handle smaller inputs, requiring reduced image resolution to maintain essential details. Preprocessing processes and resizing or downsampling CT images are crucial for CNNs to process them effectively. The input image size significantly impacts the network's efficacy, so reducing the resolution while maintaining essential characteristics and structures is essential. Downsampling methods, such as bilinear or nearest-neighbor interpolation, or more complex methods, such as wavelet-based or deep learning-based approaches, can achieve this. Within this range, The initial step in processing these images is to convert them from DICOM to PNG.

DICOM images are converted to PNG format during preprocessing through a standardized transformation process. This conversion maintains image quality. It is located on Kaggle at “https://github.com/hwanseung2/kidney-tumor-segmentation/blob/main/2D_segmentation-fastai.ipynb” (accessed on 5 March 2023). Moreover, put images in size 128×128 .

2.3. Feature Pyramid Network

Feature pyramids involve the integration of high-level feature maps spanning various scales, collaborating harmoniously with backbone networks. This synergy yields enhanced and equitable performance across multiple scales, augmenting object detection capabilities [69–71], and semantic segmentation [71–73]. The primary objective of object detection, a fundamental facet of computer vision, entails identifying the positions and categorizing objects within an image. This pursuit lays the foundation for other related tasks, including human pose estimation, expanding the scope of its utility [74]. FPN [70], Feature Pyramid Network (FPN) is a groundbreaking method in object detection using convolutional neural networks. It uses multi-scale feature data to enhance accuracy, particularly in detecting small-scale objects. The FPN’s core concept is a feature pyramid that scales features across different tiers, allowing for a more detailed representation of lower-level information. This approach surpasses the granularity of higher-level features, allowing for more accurate detection of smaller objects. FPN’s versatility makes it ideal for identifying and discerning small target objects, demonstrating its remarkable efficacy. The FPN algorithm is a trilateral process consisting of an initial traversal through a bottom-up convolutional neural network, an iterative descent from a top-down perspective, and a pivotal lateral connection that links features with their counterparts. The bottom-up phase in convolutional neural networks involves the gradual evolution of feature map dimensions as data moves forward. This dynamic fluctuation leads to the stratification of layers into discrete stages, where layers that maintain the feature map’s size coalesce into a stage. This categorization helps assign unique attributes to each extraction task, which crystallizes as the final output from each step. This process creates a composite feature pyramid, with diverse characteristics from each stage contributing to a hierarchical ensemble. The top-down progression entails an intricate movement accomplished through Upsampling, a technique consistently reliant on the interpolation principle. In this context, Upsampling orchestrates the magnification of the original image, artfully infusing new elements between pixels. This augmentation is executed via a judicious interpolation algorithm woven around the original image’s pixel configuration bedrock. It is paramount to underscore that the feature map ensuing from this Upsampling operation retains an identical dimension as the feature map is immediately contiguous in the layered hierarchy. This study compares various encoder configurations using EfficientNet, U-Net, and SE-ResNets lineages, examining their performance and result.

2.4. SE-ResNets

In SE-ResNets, add a module to residual blocks after activation, enabling access to feature maps and focusing on significant features early in the network [75]. The combination of SE block and ResNet models significantly improves image classification performance, especially for iconic and abstract images. SE blocks selectively focus on key features, acquiring more discriminative representations and improving prediction accuracy. SE-ResNet models consistently demonstrate state-of-the-art performance in benchmark datasets like ImageNet and CIFAR-10 [75]. For instance, SE-ResNet152 outperformed ResNet152 in the ImageNet dataset with a top-1 accuracy of 82.63% [75]. SE-ResNet models performed better than ResNet models on the CIFAR-10 dataset, with SE-ResNet152 achieving a test accuracy of 96.54% compared to 94.54% for ResNet152. Recommended ASC system. Note that all accuracy measurements were conducted using a four-ensemble model. The accuracy of ResNet-rev improved as the number of layers increased. Specifically, ResNet(152)-rev improved relative class performance by 12.5% compared to the baseline. Next, the classwise

accuracy of SE-ResNet-rev was evaluated and found to be superior to that of ResNet-rev at 152 layers. Finally, WaveGAN was used to augment the training data. Therefore, the class-wise accuracy of SE-ResNet(152) rev-aug was 70.5%. This was because the sampled wave may have needed to be more accurate, and the GAN-based generator and discriminator may have needed to be more adequately trained [76].

2.5. SE-ResNet + FPN Network

Using convolutional neural networks (CNNs) for semantic image attributes requires a strategic approach. CNNs, a vital tool in deep learning methodologies, follow a standardized structural composition for feature extraction. This architecture involves convolutional and pooling layers, resulting in a plain network. Examples of this genre include AlexNet and VGG. CNNs are known for their depth and intricate calculations, making them a powerful tool [77,78]. A deep-network structure for the plain network can be used to obtain a high-level feature map of each image. After downsampling, the minor details shrink or even vanish. There needs to be more data available, though, about high-resolution feature maps without downsampling. It does not allow for the easy extraction and characterization of the intricate tiny characteristics needed for our dataset. Feature pyramid network's (FPN) concept of feature pyramids served as inspiration [70]. Backbone branch: higher layers of representation amplify input characteristics crucial for discriminating and reducing unimportant variations to extract better features for the classification task [79]. Hence, we adopted SE-ResNet as our backbone to secure ample depth for rich semantic extraction due to its shortcut-connected, gradient-efficient residual mapping. The FPN framework focuses on creating a comprehensive feature pyramid with high semantic content, using a multi-pronged approach with bottom-up and top-down pathways interlinked through lateral connections.

2.6. Evaluation Metrics

F1-score and *IoU* were used in the study to evaluate the segmentation results. The Dice Similarity Coefficient (DSC), also known as the *F1-score* or Srensen-Dice index, and the Intersection-over-Union (*IoU*), also known as the Jaccard index or Jaccard similarity coefficient. The *IoU* penalizes under and over-segmentation more than the DSC, even though the DSC is defined as the harmonic mean between sensitivity and precision. The DSC is a widely used metric to assess the segmentation performance using the provided ground truth in medical images, although both scores are valid metrics [80,81]. These dimensions are expressed as follow [13]:

$$F1\text{-score} = \frac{2TP}{(TP + FP) + (TP + FN)} \quad (1)$$

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

2.7. Loss Function

The medical image predominantly comprises a minor fraction dedicated to lesion regions, exacerbating the pronounced disparity in positive and negative sample distribution. This imbalance notably impacts the accuracy of lesion region segmentation. In 2017, Tsung Yi Lin introduced focal loss as a potential remedy for this concern [82]. In object detection, focal loss was introduced to tackle the imbalanced positive and negative sample distribution challenge. In the context of medical image segmentation, where labels are influenced by clinical expertise, the potential for erroneous annotations exists. The focal loss conducts pixel-level classification, rendering the impact of these inaccuracies substantial within the network. While focal loss has its constraints, its effectiveness in mitigating imbalanced samples makes it a promising candidate for addressing similar issues in medical image segmentation. Pastor-Pellicer initially introduced the Dice loss function [83]. The Dice loss,

tailored for object segmentation utilizing the Dice similarity coefficient (DSC), emerges as a dedicated choice, particularly prevalent in medical segmentation endeavors [84]. Medical imagery often has a small region of interest, leading to a learning process skewed toward background predictions. Because the region of interest only makes up a small portion of the scan, this can affect the accurate recognition of foreground regions, frequently leading to incomplete detection or potential omissions [85].

2.8. Proposed Solution

The study introduces a methodology that utilizes the KiTS19 database to accurately segment tumors and kidneys on CT slices [67]. The method involves two primary stages: data preparation, where CT volumes are scaled and normalized to enhance data consistency and quality, and segmentation using a SE-ResNet FPN model to delineate tumors and kidneys accurately. After converting the volumes from DICOM to PNG format, the segmentation process employs a SE-ResNet FPN model to delineate the kidneys and tumor, streamlining the segmentation stage. The SE-ResNet family is an excellent choice for the encoder component, while the FPN acts as the decoder, utilizing multiscale feature maps. The method typically calculates intersection-over-union (IoU) scores for various predictions and selects the segmentation with the IoU score and F1 score to determine the optimal segmentation. The segmentation models library is an open-source resource comprising cutting-edge deep-learning models designed for image segmentation tasks, encompassing state-of-the-art architectures for semantic, instance, and panoptic segmentation, such as U-Net, PSPNet, FPN, and LinkNet. The segmentation models library offers the advantage of transfer learning and fine-tuning on custom datasets, making it a valuable tool for researchers and practitioners in image segmentation tasks. It is located on GitHub at “https://github.com/qubvel/segmentation_models” (accessed on 25 March 2023).

2.8.1. Backbone Architectures

The backbones of feature extraction networks compute image input features, and selecting the optimal network is crucial for objective task performance and DL model computational complexity. Numerous backbone networks have been designed and implemented in various DL models [86]. Further research is needed to compare feature extraction networks for Deep Learning applications [86]. This study evaluates five existing feature extraction backbone networks for a single model to determine the most effective combination. Unsuitable backbones can degrade performance, be computationally expensive, and be complex. The study aims to improve model performance and reduce computational costs. Figure 2 summarizes the categories (family) and names of the backbones.

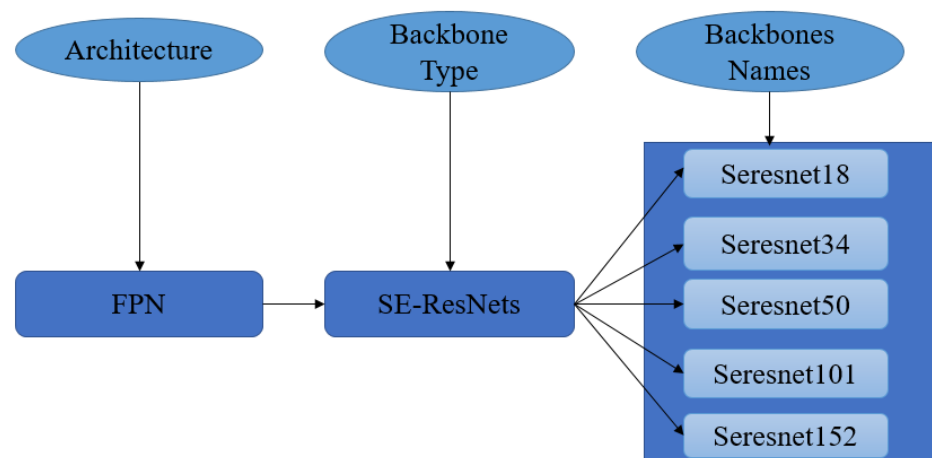


Figure 2. Architecture, Backbone Type and Backbones Names.

2.8.2. Methodology

In this work, semantic segmentation of renal tumors Combines a semantic segmentation model and five feature extraction networks, which results in the proposed methodology. This modular design aims to determine the optimal segmentation solution for kidney tumors. Additionally, The optimal Loss function is investigated and evaluated from both model and backbone perspectives, and the methodology is proposed, as shown in Figure 3.

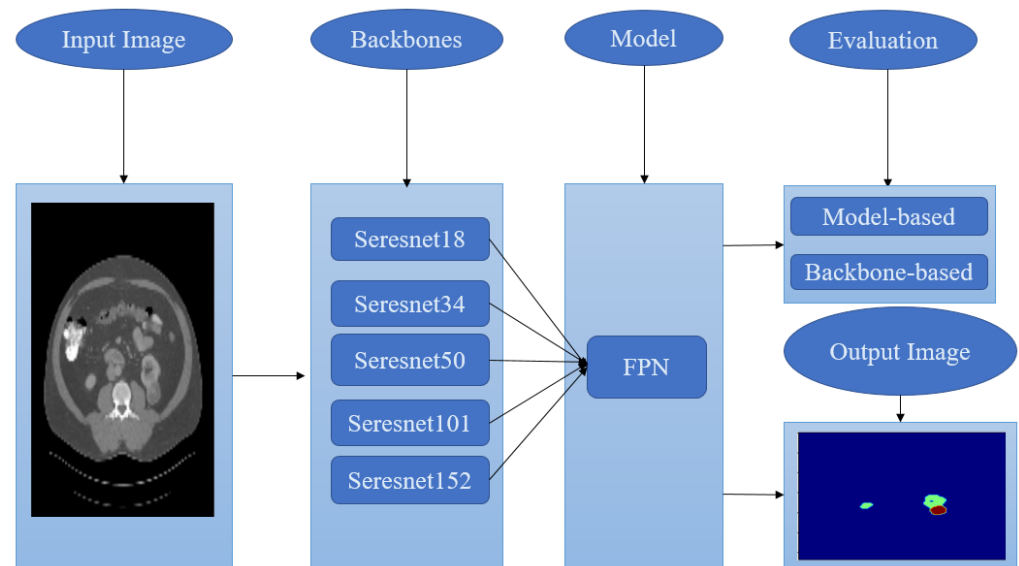


Figure 3. Proposed methodology illustration.

The proposed segmentation architecture combines a semantic segmentation model with five feature extraction networks, resulting in output segmentation images and numerical experimental results.

2.9. Implementation

In this paper, we implemented all algorithms using Python 3.9.16 on a DELL CORE i5 personal computer with the support of Anaconda and Jupyter. Initially sized at 512×512 pixels, the original images underwent uniform resizing to 128×128 pixels, serving as the segmentation model's input. Focus in this section revolves around constructing five models from the SE-ResNets family, namely, 'Seresnet18', 'Seresnet34', 'Seresnet50', 'Seresnet101' and 'Seresnet152', employing the FPN (Feature Pyramid Network) architecture. This approach aims to leverage the capabilities of the SE-ResNets and the benefits of the FPN framework to enhance the accuracy and efficiency of our segmentation model.

2.10. Model Setup

Every model undergoes training for a total of 50 epochs. In order to prevent overfitting, the training process concludes when the validation loss stays at or above 0.0001. All backbone architectures are initialized with weights pre-trained on ImageNet to expedite convergence. The specifics of baseline model configurations and hyperparameters for all deep learning models are provided in Table 2.

Table 2. Models' hyperparameter setup.

Hyperparameter	Settings
Activation	'softmax'
Optimizer	Adam
Loss function	Focal Loss, Dice Loss
Learning rate	0.0001
Batch size	32
Epochs	50
Metrics	IoU, F1-score
Input images size	128 × 128

3. Results and Discussion

This section discusses assembling the five models, Seresnet18, Seresnet34, Seresnet50, Seresnet101, and Seresnet152, using FPN architecture. To train all models using SE-ResNet FPN architecture.

3.1. Experimental Results

In Table 3, the accuracy of the ensemble model of Seresnet50 for Background, Kidney, Kidney Tumor, and Mean IoU outcomes is presented.

Table 3. Outcomes of the Ensemble Model of Seresnet50.

IoU Scores	Obtained Values
Background IoU Score	0.999
Kidney IoU Score	0.980
Tumor IoU Score	0.984
Mean IoU Score	0.988

The accuracy of Seresnet50 is assessed through IoU scores, which gauge the concordance between predicted and actual segments. Elevated scores signify strong correspondence between predictions and actual structures in medical images, indicative of proficient identification and segmentation. High IoU scores, for instance, Background 0.999, Kidney 0.980, and Tumor 0.984 underscore the model's precision in outlining these specific areas.

3.2. Discussion

3.2.1. Result Analysis

ImageNet assessed and optimized best-performing models, despite patch-wise classification using local features [48]. This research investigates the architecture's ability to extract global features from CT images of renal cancer and classify unseen images, revealing that extracting nuclei and tissue organization features is more valuable than nuclei-scale features for deciphering image classes.

Evaluated different models based on their performance and employed a metric called Mean IoU (Intersection over Union) to assess the accuracy of the models. The study compared several models based on the SE-ResNet architecture with FPN (Feature Pyramid Network) and presented the results in Table 4. Each model was evaluated regarding background IoU score, kidney IoU score, tumor IoU score, and mean IoU score. The IoU scores indicate how well the models performed in segmenting different regions of the CT images. A higher IoU score signifies better performance, indicating higher overlap between the predicted and ground truth regions. The results demonstrate that all the SE-ResNet models achieved excellent IoU scores across the different categories. For example, the 'Seresnet18' model achieved a background IoU score of 0.999, a kidney IoU score of 0.972, a tumor IoU score of 0.971, and a mean IoU score of 0.981. Similarly, the other models, 'Seresnet34', 'Seresnet50', 'Seresnet101' and 'Seresnet152,' also achieved high IoU scores across the categories. Based on these results, it can be inferred that the SE-ResNet models

with FPN architecture effectively extracted global features from CT images of renal cancer and utilized them to classify unseen images accurately. Figure 4 compares results for the SE-ResNet with FPN Architectures Using Mean IoU.

Table 4. Results for the SE-ResNet with FPN Architectures Using Mean IoU.

SE-ResNet	Background	Kidney	Tumor	Mean
Seresnet18	0.999	0.972	0.971	0.981
Seresnet34	0.999	0.972	0.971	0.981
Seresnet50	0.999	0.980	0.984	0.988
Seresnet101	0.999	0.976	0.982	0.986
Seresnet152	0.999	0.979	0.984	0.987

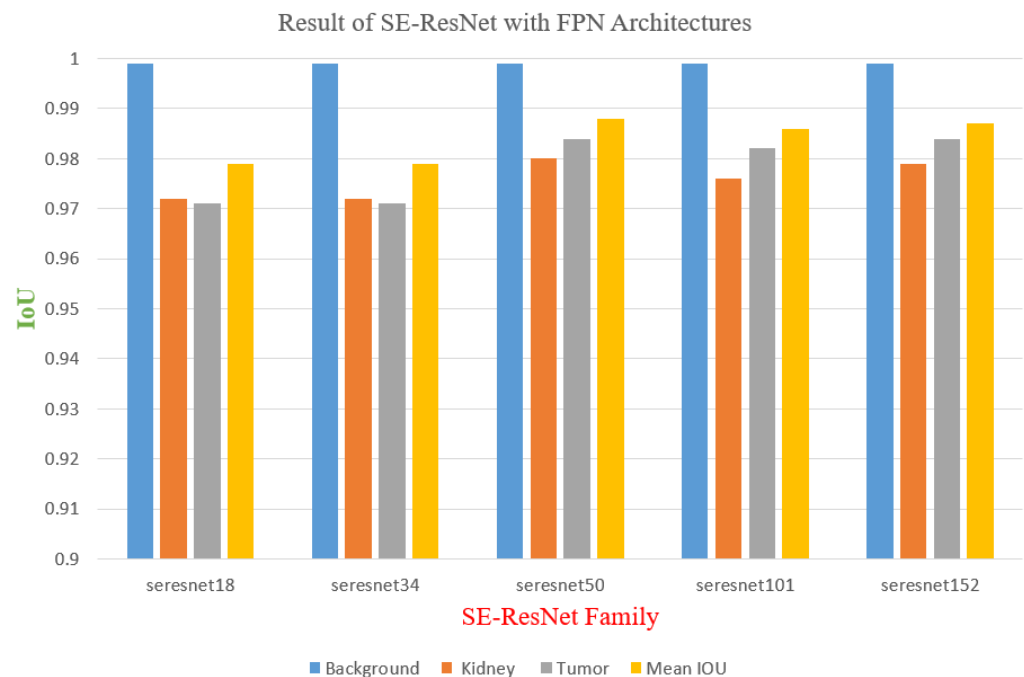


Figure 4. Result of SE-ResNet with FPN Architectures.

Table 5 presents the results of the SE-ResNet models in terms of accuracy and loss. IoU score (Intersection over Union) and F1-score using validation images for evaluation. These metrics provide insights into the performance of the models in terms of overall accuracy and their ability to accurately segment and classify objects in the images. The accuracy represents the percentage of correctly classified instances from the total instances in the validation dataset. In this case, the accuracy is measured by the loss value, which is inversely proportional to accuracy. Therefore, a lower loss value indicates higher accuracy. For all the SE-ResNet models, including 'Seresnet18', 'Seresnet34', 'Seresnet50', 'Seresnet101', and 'Seresnet152' the loss values are around 0.752, indicating good accuracy in the range of 75.2%. The IoU score, the Jaccard Index, measures the overlap between the predicted and ground truth segmentation. A higher IoU score indicates better performance in accurately identifying and localizing objects. The SE-ResNet models achieved high IoU scores, ranging from 0.979 to 0.987. For example, the 'Seresnet18' and 'Seresnet34' models achieved an IoU score of 0.979, while 'Seresnet50' achieved a slightly higher score of 0.987. Similarly, 'Seresnet101' and 'Seresnet152' achieved IoU scores of 0.985 and 0.986, respectively. These scores suggest that the models effectively segment the objects of interest in the validation images. The F1-score measures the models' accuracy in terms of precision and recall. It considers both false positives and false negatives in the predictions. The SE-ResNet models achieved high F1-scores, ranging from 0.989 to 0.993. The scores indicate a high level of precision and recall in classifying the objects in the images, with 'Seresnet50', 'Seresnet101',

and ‘Seresnet152’ achieving slightly higher scores compared to ‘Seresnet18’ and ‘Seresnet34’. These results demonstrate that the SE-ResNet models performed well in terms of accuracy, IoU score, and F1 score when evaluated on the validation images. They showcase the models’ ability to classify accurately and segment objects, indicating their potential usefulness in tasks such as image classification and object detection. SE-ResNet accuracy, IoU, F1-score, and loss results are shown in Figure 5.

Table 5. Results for the SE-ResNet Accuracy IoU and F1-score Using Validation Images.

SE-ResNet	Loss	IoU Score	F1-Score
Seresnet18	0.753	0.979	0.989
Seresnet34	0.753	0.979	0.989
Seresnet50	0.752	0.987	0.993
Seresnet101	0.752	0.985	0.992
Seresnet152	0.752	0.986	0.993

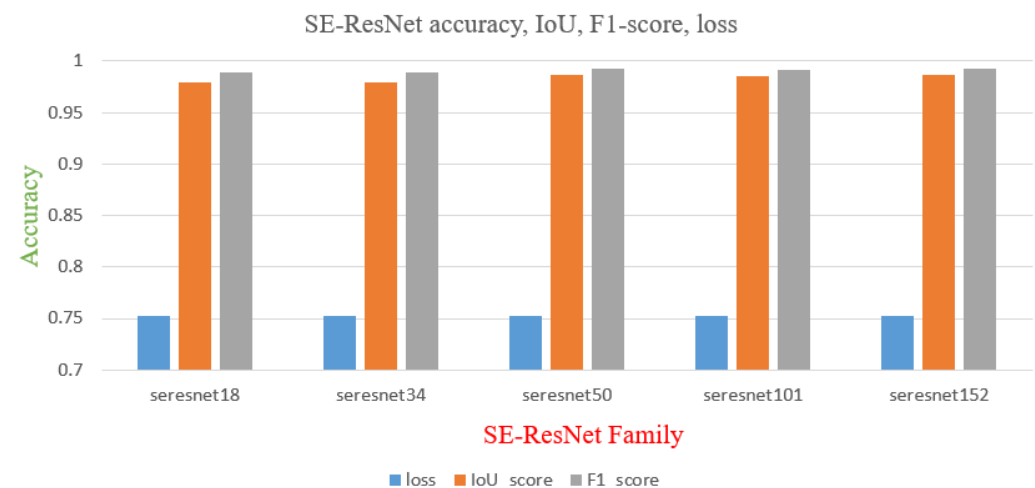


Figure 5. SE-ResNet Accuracy IoU and F1-score Using Validation Images.

3.2.2. Model Accuracy and Model Loss

The SE-ResNet model’s performance is evaluated using metrics like loss and IoU Score to minimize discrepancies between predicted and actual values. It measures the overlap between predicted bounding boxes and the ground truth, assessing accuracy in segmenting kidney tumors and kidneys. IoU Score measures the overlap between predicted and accurate segmentation, while the loss function quantifies the model’s performance in minimizing the error between predicted outputs and ground truth, optimizing segmentation results. In Figures 6 and 7, Model accuracy and Model loss Seresnet50. Our strategies remained immune to overfitting issues due to the appropriate dataset.

The discussion on the performance evaluation of different versions of the SE-ResNet model in terms of loss and IoU score. Loss is a metric that measures how well the model minimizes the difference between predicted and actual values. IoU score, on the other hand, measures the overlap between the predicted bounding boxes and the ground truth bounding boxes. According to the results, the models exhibit consistent performance in terms of loss, with Seresnet18 and Seresnet34 having a loss value of 0.753. The other models, namely Seresnet50, Seresnet101, and Seresnet152, show slightly lower loss values. Lower loss values indicate that these models are better at minimizing the difference between their predictions and the actual values. However, when it comes to the IoU score, there is some variation among the models. Seresnet50 achieves the highest IoU score of 0.987, indicating that it is the most accurate model for object detection.

On the other hand, Seresnet18 and Seresnet34 have the lowest IoU score of 0.979, suggesting that they may struggle to detect objects accurately in some cases. Overall,

the SE-ResNet models perform well in object detection tasks, with Seresnet50 being the most accurate model based on the IoU score. However, it is significant to consider other factors, such as model size and computational complexity, when selecting the most appropriate model for a specific task.

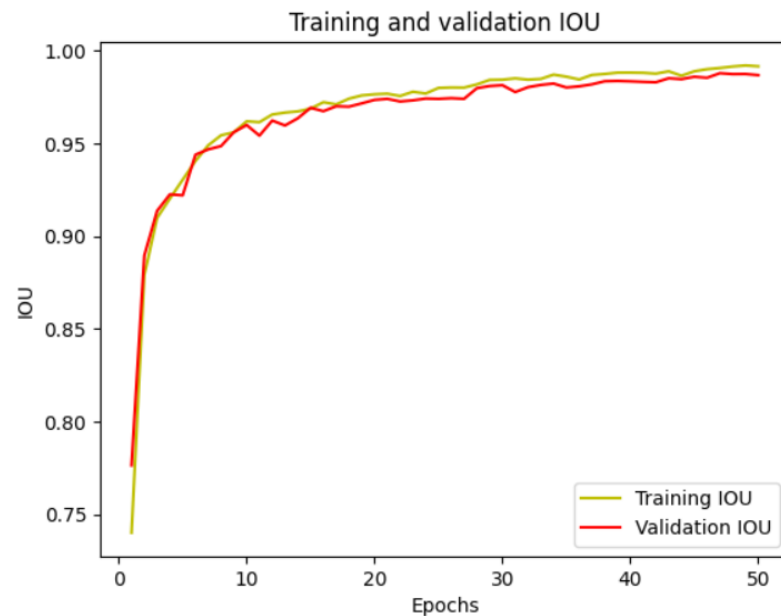


Figure 6. Model IoU for Seresnet50.

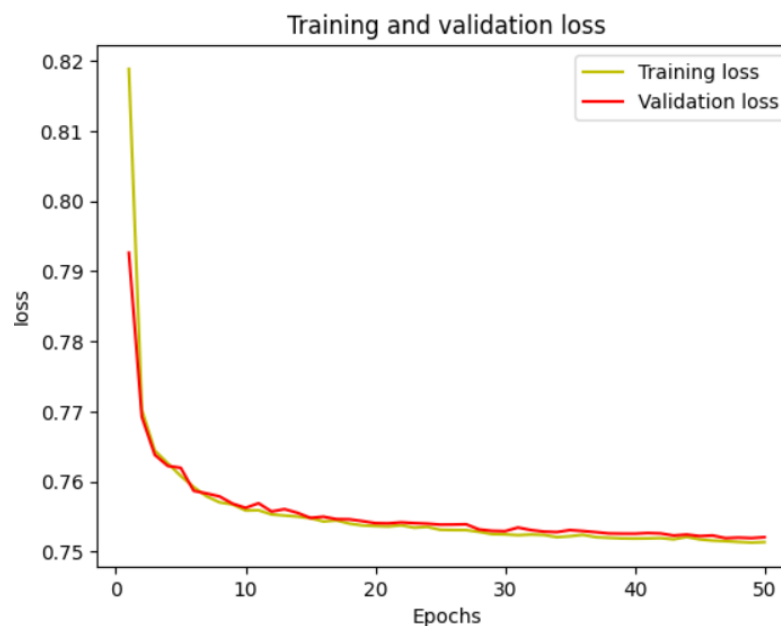


Figure 7. Model Loss for Seresnet50.

3.2.3. Prediction on Validation Images

Figure 8 presents an illustrative example of kidney and kidney tumor segmentation using the Seresnet18 and Seresnet34 models. The annotation images showcase the annotations of the kidney and tumor regions, represented in green and blue colors. These annotations serve as the ground truth for evaluating the segmentation model's performance. In contrast, the result images exhibit the model's kidney and tumor segmentation predictions. The predicted kidney region is green, while the tumor region is brown. The Seresnet18 and Seresnet34 model's segmentation outputs are visually depicted in these images, clearly

representing their performance in identifying and delineating kidney and tumor regions within CT scans.

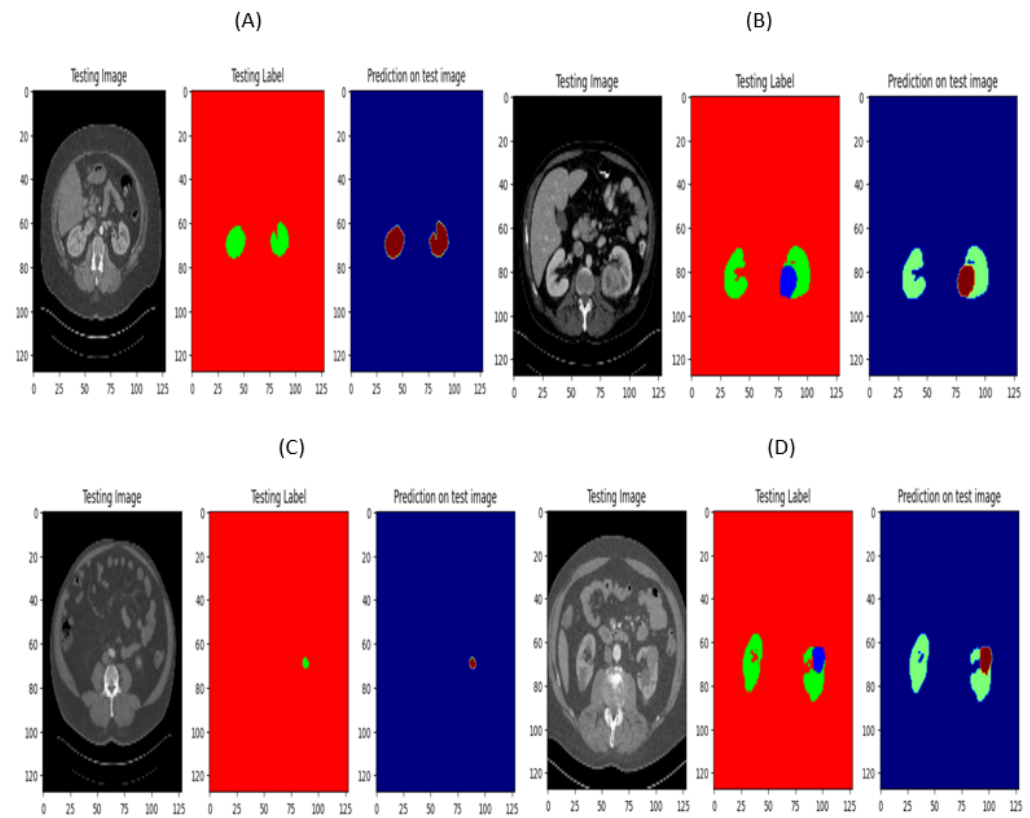


Figure 8. (A) seresnet18 image normal and (B) seresnet18 image with tumor (C) seresnet34 image normal and (D) seresnet34 image with tumor.

3.2.4. Prediction on Testing Images

An example of segmenting the kidney and a kidney tumor utilizing the Seresnet50, Seresnet101, and Seresnet152 models is shown in Figure 9 using test images. The annotation images annotate the kidney and tumour regions in green and brown. The performance of the segmentation model is assessed using these annotations as the source of truth. On the other hand, the segmentation predictions for the kidney and tumor in the outcome images are visible. These photos clearly show the segmentation outputs of the Seresnet50, Seresnet101, and Seresnet152 models, demonstrating their effectiveness in locating and defining kidney and tumor regions inside CT scans.

Table 6 includes various architectures, such as U-Net, FPN, and others. Each architecture is associated with its respective kidney and tumor detection scores. Based on the results, our approach utilizing the FPN (Feature Pyramid Network) with different versions of the SE-ResNet model (Seresnet18, Seresnet34, Seresnet50, Seresnet101, and Seresnet152) achieves high scores for both kidney and tumor detection. FPN-Seresnet50 demonstrates the highest performance, with a kidney score of 0.980 and a tumor score of 0.984. As a result, it can be concluded that our kidney and tumor detection method utilizing FPN-Seresnet50 is exact and efficient. Furthermore, as an additional enhancement, we incorporated ResNet50 into our approach. The outcomes yielded by this addition were marginally lower than the results attained with SeresNet50. It is important to note that different architectures show varying performance across the tasks. The scores in the table provide a comparative analysis of the different approaches used in kidney and tumor detection. Figure 10 compares different segmentation approaches.

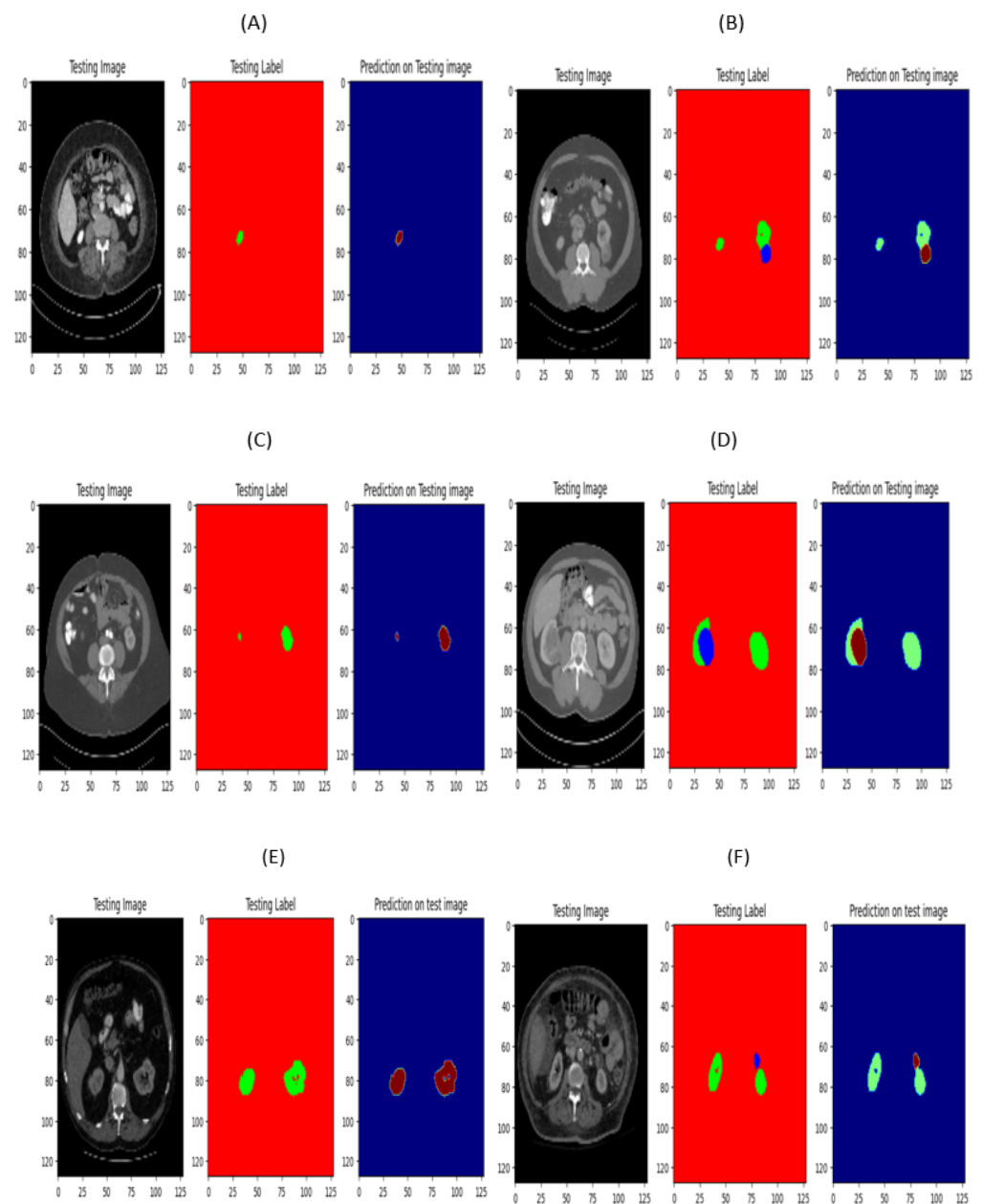


Figure 9. (A) seresnet50 image normal and (B) seresnet50 image with tumor. (C) seresnet101 image normal and (D) seresnet101 image with tumor. (E) seresnet152 image normal and (F) seresnet152 image with tumor.

Table 6. Evaluating Pretrained Architecture Approaches Against Prior Methods for KiTS19 Dataset Segmentation.

Reference	Architecture	Kidney Dice	Tumor Dice	Kidney IoU Scores	Tumor IoU Scores
[87]	nnU-Net	0.969	0.919	-	-
[87]	Hybrid V-Net	0.962	0.913	-	-
[87]	U-Net3+	0.959	0.909	-	-
[87]	SRE Net 2D-CNN	0.979	0.925	-	-
[88]	FR2PAttU-Net	0.948	0.911	-	-
[89]	U-Net EfficientNet-B0	0.984	0.980	0.969	0.960
[89]	U-Net EfficientNet-B1	0.981	0.980	0.963	0.961
[89]	U-Net EfficientNet-B2	0.983	0.982	0.966	0.965
[89]	U-Net EfficientNet-B3	0.985	0.980	0.970	0.960

Table 6. Cont.

Reference	Architecture	Kidney Dice	Tumor Dice	Kidney IoU Scores	Tumor IoU Scores
[89]	U-Net EfficientNet-B4	0.987	0.984	0.974	0.968
[89]	U-Net EfficientNet-B5	0.986	0.983	0.972	0.966
[89]	U-Net EfficientNet-B6	0.985	0.981	0.971	0.963
[89]	U-Net EfficientNet-B7	0.988	0.981	0.977	0.962
Proposed	FPN- Seresnet18	0.986	0.985	0.972	0.971
Proposed	FPN- Seresnet34	0.986	0.985	0.972	0.971
Proposed	FPN- Seresnet50	0.990	0.992	0.980	0.984
Proposed	FPN- Seresnet101	0.988	0.991	0.976	0.982
Proposed	FPN- Seresnet152	0.989	0.992	0.979	0.984
Proposed	FPN- Resnet50	0.987	0.988	0.974	0.977

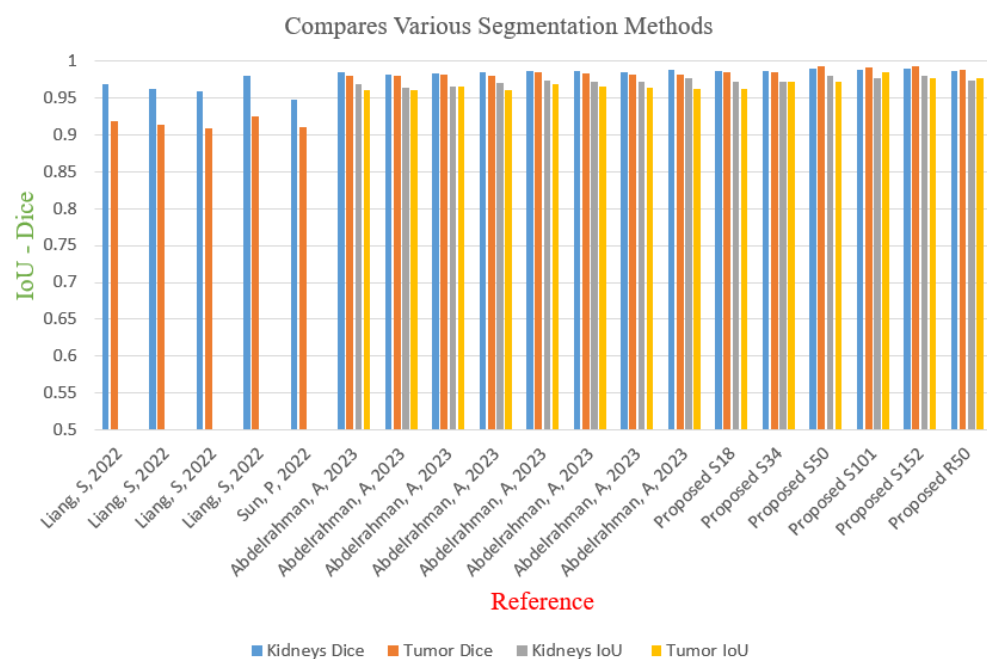


Figure 10. Comparing Pretrained Architecture Approaches against Prior Methods for KiTS19 Dataset Segmentation Evaluation [87–89].

3.2.5. Computational Efficiency and Training Time

The research analyzes EfficientNetB4 [89], EfficientNetB7 [89], and Seresnet50 in terms of computational requirements. EfficientNetB4 has a smaller model size, while EfficientNetB7 and Seresnet50 have more complex architectures. Seresnet50's complexity leads to longer training periods, highlighting the interplay between model structure and training efficiency. These models' computational expenses and performance can vary based on implementation, hardware setups, and dataset characteristics. Table 7 summarizes parameter counts and time-related costs for each model.

Table 7. Computational efficiency and training time.

Reference	Total Params	Trainable Params	Non-Trainable Params	Time per Second
EfficientNetB4 [89]	25,735,307	25,608,123	127,184	114,386
EfficientNetB7 [89]	75,048,387	74,735,682	312,704	118,455
Proposed Seresnet50	29,460,211	29,404,787	55,424	120,970

4. Conclusions

In conclusion, using a SE-ResNet FPN model, this paper proposes a method for segmenting tumors and kidneys on CT slices. The proposed method comprises data preparation, segmentation, and an ensemble phase. Classifying CT images necessitates the preprocessing, downsampling, and resizing the images. In addition, the IoU and F1-score metrics are discussed as common evaluation metrics for image segmentation tasks. The paper concludes with a summary of related techniques in the literature. The proposed method can increase segmentation accuracy and kidney and tumor diagnosis on CT slices.

The study investigated the performance of the Feature Pyramid Network (FPN) architecture in combination with different models from the SE-ResNet family for image segmentation, specifically in kidney cancer CT images. The findings indicated that the models fine-tuned on ImageNet surpassed patch-wise classification approaches regarding accuracy and sensitivity. The SE-ResNet models exhibited impressive results in segmenting the background, kidney, and tumor regions, as evidenced by mean IoU (Intersection over Union) scores ranging from 0.981 to 0.988. Seresnet50 demonstrated the highest IoU score of 0.988, indicating its effectiveness in accurately segmenting tumors. Comparison with prior studies further highlighted the superior performance of the models employed in this research, as they achieved higher IoU scores for kidney and tumor segmentation. These outcomes emphasize that combining the FPN architecture and SE-ResNet models can discern and segment the relevant regions of interest in kidney cancer CT images. Overall, this study underscores the potential of utilizing deep learning approaches, specifically the FPN architecture and SE-ResNet models, for precise and accurate image segmentation tasks in the context of kidney cancer histology analysis.

Also, for the Dice similarity coefficient (DSC) Scores, Table 6 showcases the remarkable performance of the SE-ResNet model when dealing with the specific image segmentation tasks examined in this study. The models achieved impressive Dice Scores for background, kidney, and tumor segmentation, with mean Dice Scores ranging from 0.989 for Seresnet18 and Seresnet34. Notably, the performance generally improved as the SE-ResNet architecture version progressed from Seresnet18 to Seresnet152. For instance, the Dice Score increased from 0.989 for Seresnet18 to 0.993 for Seresnet50 and Seresnet152, then slightly declined for Seresnet101 to 0.992. The background segmentation Dice Scores were consistently exceptionally high, with a score of 0.999, across all SE-ResNet designs (Seresnet18-Seresnet152). Seresnet50 achieved the highest Dice Score of 0.990 for segmenting kidneys, while Seresnet50 and Seresnet152 obtained the top Dice Score of 0.992 for segmenting tumors. These findings demonstrate that the SE-ResNet designs effectively address the challenges posed by these segmentation tasks.

The proposed strategy was evaluated on the KiTS19 benchmark dataset for kidney tumor segmentation. It would be intriguing to see how well the method performs on other medical imaging datasets, such as lung, brain, and liver imaging. Comparing the performance of the proposed method to that of other cutting-edge techniques on various datasets could help establish its generalizability and robustness. Exploration of other deep learning architectures: For image segmentation tasks, the authors of this paper used the FPN architecture in conjunction with SE-ResNet models. Nonetheless, many other deep learning architectures, including Attention U-Net, DeepLab, and Mask R-CNN, could be investigated for this purpose. Future research could compare the performance of various architectures to the performance of the proposed method.

Author Contributions: Conceptualization, A.A. and S.V.; Methodology, A.A.; Software, A.A.; Formal analysis, S.V.; Investigation, A.A. and S.V.; Writing—original draft, A.A.; Writing—review & editing, S.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhou, X.; Menche, J.; Barabási, A.L.; Sharma, A. Human Symptoms-Disease Network. *Nat. Commun.* **2014**, *5*, 4212. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Rasmussen, J. Diagnostic Reasoning in Action. *IEEE Trans. Syst. Man Cybern.* **1993**, *23*, 981–992. [\[CrossRef\]](#)
3. Clark, L.A.; Cuthbert, B.; Lewis-Fernández, R.; Narrow, W.E.; Reed, G.M. Three Approaches to Understanding and Classifying Mental Disorder: ICD-11, DSM-5, and the National Institute of Mental Health’s Research Domain Criteria (RDoC). *Psychol. Sci. Public Interes.* **2017**, *18*, 72–145. [\[CrossRef\]](#)
4. Dalby, W. Section of Otolaryngology. *Br. Med. J.* **1895**, *2*, 1289–1294. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Scheuermann, R.H.; Ceusters, W.; Smith, B. Toward an Ontological Treatment of Disease and Diagnosis Department of Pathology and Division of Biomedical Informatics, University of Texas. *AMIA Summit Transl. Bioinform.* **2009**, *2009*, 116–120.
6. Croft, P.; Altman, D.G.; Deeks, J.J.; Dunn, K.M.; Hay, A.D.; Hemingway, H.; LeResche, L.; Peat, G.; Perel, P.; Petersen, S.E.; et al. The Science of Clinical Practice: Disease Diagnosis or Patient Prognosis? Evidence about “What Is Likely to Happen” Should Shape Clinical Practice. *BMC Med.* **2015**, *13*, 1–8. [\[CrossRef\]](#)
7. Torres, A.; Nieto, J.J. Fuzzy Logic in Medicine and Bioinformatics. *J. Biomed. Biotechnol.* **2006**, *2006*, 1–7. [\[CrossRef\]](#)
8. Alam, R.; Cheraghi-Sohi, S.; Panagioti, M.; Esmail, A.; Campbell, S.; Panagopoulou, E. Managing Diagnostic Uncertainty in Primary Care: A Systematic Critical Review. *BMC Fam. Pract.* **2017**, *18*, 79. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Malmir, B.; Amini, M.; Chang, S.I. A Medical Decision Support System for Disease Diagnosis under Uncertainty. *Expert Syst. Appl.* **2017**, *88*, 95–108. [\[CrossRef\]](#)
10. Nilashi, M.; Ibrahim, O.B.; Ahmadi, H.; Shahmoradi, L. An Analytical Method for Diseases Prediction Using Machine Learning Techniques. *Comput. Chem. Eng.* **2017**, *106*, 212–223. [\[CrossRef\]](#)
11. Nilashi, M.; Ibrahim, O.; Ahmadi, H.; Shahmoradi, L.; Farahmand, M. A Hybrid Intelligent System for the Prediction of Parkinson’s Disease Progression Using Machine Learning Techniques. *Biocybern. Biomed. Eng.* **2018**, *38*, 1–15. [\[CrossRef\]](#)
12. Nilashi, M.; Ibrahim, O.; Dalvi, M.; Ahmadi, H.; Shahmoradi, L. Accuracy Improvement for Diabetes Disease Classification: A Case on a Public Medical Dataset. *Fuzzy Inf. Eng.* **2017**, *9*, 345–357. [\[CrossRef\]](#)
13. Abdelrahman, A.; Viriri, S. Kidney Tumor Semantic Segmentation Using Deep Learning: A Survey of State-of-the-Art. *J. Imaging* **2022**, *8*, 55. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Salih, O.; Duffy, K.J. Optimization Convolutional Neural Network for Automatic Skin Lesion Diagnosis Using a Genetic Algorithm. *Appl. Sci.* **2023**, *13*, 3248. [\[CrossRef\]](#)
15. Gur, D.; Sumkin, J.H.; Rockette, H.E.; Ganott, M.; Hakim, C.; Hardesty, L.; Poller, W.R.; Shah, R.; Wallace, L. Changes in Breast Cancer Detection and Mammography Recall Rates after the Introduction of a Computer-Aided Detection System. *J. Natl. Cancer Inst.* **2004**, *96*, 185–190. [\[CrossRef\]](#)
16. Destounis, S.V.; DiNitto, P.; Logan-Young, W.; Bonaccio, E.; Zuley, M.L.; Willison, K.M. Can Computer-Aided Detection with Double Reading of Screening Mammograms Help Decrease the False-Negative Rate? Initial Experience. *Radiology* **2004**, *232*, 578–584. [\[CrossRef\]](#)
17. Doi, K. Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential. *Comput. Med. Imaging Graph.* **2007**, *31*, 198–211. [\[CrossRef\]](#)
18. Salih, O.; Viriri, S. Skin Lesion Segmentation Using Stochastic Region-Merging and Pixel-Based Markov Random Field. *Symmetry* **2020**, *12*, 1224. [\[CrossRef\]](#)
19. Li, Q.; Li, F.; Suzuki, K.; Shiraishi, J.; Abe, H.; Engelmann, R.; Nie, Y.; MacMahon, H.; Doi, K. Computer-Aided Diagnosis in Thoracic CT. *Semin. Ultrasound CT MRI* **2005**, *26*, 357–363. [\[CrossRef\]](#)
20. Liu, Y.Y.; Huang, Z.H.; Huang, K.W. Deep Learning Model for Computer-Aided Diagnosis of Urolithiasis Detection from Kidney–Ureter–Bladder Images. *Bioengineering* **2022**, *9*, 811. [\[CrossRef\]](#)
21. Panayides, A.S.; Amini, A.; Filipovic, N.D.; Sharma, A.; Tsaftaris, S.A.; Young, A.; Foran, D.; Do, N.; Golemati, S.; Kurc, T.; et al. AI in Medical Imaging Informatics: Current Challenges and Future Directions. *IEEE J. Biomed. Heal. Inf.* **2020**, *24*, 1837–1857. [\[CrossRef\]](#)
22. Lim, E.J.; Castellani, D.; So, W.Z.; Fong, K.Y.; Li, J.Q.; Tiong, H.Y.; Gadzhiev, N.; Heng, C.T.; Teoh, J.Y.C.; Naik, N.; et al. Radiomics in Urolithiasis: Systematic Review of Current Applications, Limitations, and Future Directions. *J. Clin. Med.* **2022**, *11*, 5151. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Vishwanath, V.; Jafarieh, S.; Rembielak, A. The Role of Imaging in Head and Neck Cancer: An Overview of Different Imaging Modalities in Primary Diagnosis and Staging of the Disease. *J. Contemp. Brachyther.* **2020**, *12*, 512–518. [\[CrossRef\]](#)
24. Bi, W.L.; Hosny, A.; Schabath, M.B.; Giger, M.L.; Birkbak, N.J.; Mehrtash, A.; Allison, T.; Arnaout, O.; Abbosh, C.; Dunn, I.F.; et al. Artificial Intelligence in Cancer Imaging: Clinical Challenges and Applications. *CA Cancer J. Clin.* **2019**, *69*, 127–157. [\[CrossRef\]](#) [\[PubMed\]](#)

25. Salmanpour, M.R.; Shamsaei, M.; Saberi, A.; Setayeshi, S.; Klyuzhin, I.S.; Sossi, V.; Rahmim, A. Optimized Machine Learning Methods for Prediction of Cognitive Outcome in Parkinson's Disease. *Comput. Biol. Med.* **2019**, *111*, 103347. [\[CrossRef\]](#) [\[PubMed\]](#)
26. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-Based Learning Applied to Document Recognition. *Proc. IEEE* **1998**, *86*, 2278–2323. [\[CrossRef\]](#)
27. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A Survey on Deep Learning in Medical Image Analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [\[CrossRef\]](#)
28. Sarvamangala, D.R.; Kulkarni, R.V. Convolutional Neural Networks in Medical Image Understanding: A Survey. *Evol. Intell.* **2022**, *15*, 1–22. [\[CrossRef\]](#)
29. Chan, H.P.; Hadjiiski, L.M.; Samala, R.K. Computer-Aided Diagnosis in the Era of Deep Learning. *Med. Phys.* **2020**, *47*, e218–e227. [\[CrossRef\]](#)
30. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 6999–7019. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Shi, Y.; Yang, W.; Gao, Y.; Shen, D. Does Manual Delineation Only Provide the Side Information in CT Prostate Segmentation? *Lect. Notes Comput. Sci.* **2017**, *10435*, 692–700.
32. He, B.; Xiao, D.; Hu, Q.; Jia, F. Automatic Magnetic Resonance Image Prostate Segmentation Based on Adaptive Feature Learning Probability Boosting Tree Initialization and CNN-ASM Refinement. *IEEE Access* **2017**, *6*, 2005–2015. [\[CrossRef\]](#)
33. Mortazi, A.; Karim, R.; Rhode, K.; Burt, J.; Bagci, U. CardiacNET: Segmentation of Left Atrium and Proximal Pulmonary Veins from MRI Using Multi-View CNN. *Lect. Notes Comput. Sci.* **2017**, *10434*, 377–385.
34. Patravali, J.; Jain, S.; Chilamkurthy, S. 2D-3D Fully Convolutional Neural Networks for Cardiac MR Segmentation. *Lect. Notes Comput. Sci.* **2018**, *10663*, 130–139.
35. Moeskops, P.; Viergever, M.A.; Mendrik, A.M.; De Vries, L.S.; Benders, M.J.N.L.; Isgum, I. Automatic Segmentation of MR Brain Images with a Convolutional Neural Network. *IEEE Trans. Med. Imaging* **2016**, *35*, 1252–1261. [\[CrossRef\]](#)
36. Wang, G.; Li, W.; Zuluaga, M.A.; Pratt, R.; Patel, P.A.; Aertsen, M.; Doel, T.; David, A.L.; Deprest, J.; Ourselin, S.; Vercauteren, T. Interactive Medical Image Segmentation Using Deep Learning with Image-Specific Fine Tuning. *IEEE Trans. Med. Imaging* **2018**, *37*, 1562–1573. [\[CrossRef\]](#)
37. Salmanpour, M.R.; Hosseinzadeh, M.; Masoud, S. Computer Methods and Programs in Biomedicine Fusion-Based Tensor Radiomics Using Reproducible Features: Application to Survival Prediction in Head and Neck Cancer. *Comput. Methods Programs Biomed.* **2023**, *240*, 107714. [\[CrossRef\]](#)
38. Salmanpour, M.R.; Rezaei, S.M.; Hosseinzadeh, M.; Rahmim, A. Deep versus Handcrafted Tensor Radiomics Features: Prediction of Survival in Head and Neck Cancer Using Machine Learning and Fusion Techniques. *Diagnostics* **2023**, *13*, 1696. [\[CrossRef\]](#)
39. Jahangirimehr, A.; Abdolahi Shahvali, E.; Rezaei, S.M.; Khalighi, A.; Honarmandpour, A.; Honarmandpour, F.; Labibzadeh, M.; Bahmanyari, N.; Heydarheydari, S. Machine Learning Approach for Automated Predicting of COVID-19 Severity Based on Clinical and Paraclinical Characteristics: Serum Levels of Zinc, Calcium, and Vitamin D. *Clin. Nutr. ESPEN* **2022**, *51*, 404–411. [\[CrossRef\]](#)
40. Lee, K.; Zung, J.; Li, P.; Jain, V.; Seung, H.S. Superhuman Accuracy on the SNEMI3D Connectomics Challenge. *arXiv* **2017**, arXiv:1706.00120.
41. Isensee, F.; Petersen, J.; Klein, A.; Zimmerer, D.; Jaeger, P.F.; Kohl, S.; Wasserthal, J.; Koehler, G.; Norajitra, T.; Wirkert, S.; Maier-Hein, K.H. nnU-Net: Self-Adapting Framework for U-Net-Based Medical Image Segmentation. *Inform. Aktuell* **2019**, *22*. [\[CrossRef\]](#)
42. Vu, M.H.; Grimbergen, G.; Nyholm, T.; Löfstedt, T. Evaluation of Multislice Inputs to Convolutional Neural Networks for Medical Image Segmentation. *Med. Phys.* **2020**, *47*, 6216–6231. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Motzer, R.J.; Bander, N.H.; Nanus, D. M. Renal-Cell Carcinoma. *N. Engl. J. Med.* **1996**, *335*, 865–875. [\[CrossRef\]](#)
44. Liu, J.; Yildirim, O.; Akin, O.; Tian, Y. AI-Driven Robust Kidney and Renal Mass Segmentation and Classification on 3D CT Images. *Bioengineering* **2023**, *10*, 116. [\[CrossRef\]](#)
45. Shehata, M.; Abouelkheir, R.T.; Gayhart, M.; Van Bogaert, E.; Abou El-Ghar, M.; Dwyer, A.C.; Ouseph, R.; Yousaf, J.; Ghazal, M.; Contractor, S.; et al. Role of AI and Radiomic Markers in Early Diagnosis of Renal Cancer and Clinical Outcome Prediction: A Brief Review. *Cancers* **2023**, *15*, 2835. [\[CrossRef\]](#)
46. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer Statistics, 2015. *CA Cancer J. Clin.* **2015**, *65*, 5–29. [\[CrossRef\]](#)
47. Muglia, V.F.; Prando, A. Renal Cell Carcinoma: Histological Classification and Correlation with Imaging Findings. *Radiol. Bras.* **2015**, *48*, 166–174. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Humphrey, P.A.; Moch, H.; Cubilla, A.L.; Ulbright, T.M.; Reuter, V.E. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs—Part B: Prostate and Bladder Tumours. *Eur. Urol.* **2016**, *70*, 106–119. [\[CrossRef\]](#)
49. Rendon, R.A. Active Surveillance as the Preferred Management Option for Small Renal Masses. *J. Can. Urol. Assoc.* **2010**, *4*, 136–138. [\[CrossRef\]](#)
50. Mindrup, S.R.; Pierre, J.S.; Dahmouch, L.; Konety, B.R. The Prevalence of Renal Cell Carcinoma Diagnosed at Autopsy. *BJU Int.* **2005**, *95*, 31–33. [\[CrossRef\]](#)
51. Nadimi-Shahraki, M.H.; Taghian, S.; Mirjalili, S.; Abualigah, L. Binary Aquila Optimizer for Selecting Effective Features from Medical Data: A COVID-19 Case Study. *Mathematics* **2022**, *10*, 1929. [\[CrossRef\]](#)

52. Rezaeijoo, S.M.; Nesheli, S.J.; Serj, M.F.; Birgani, M.J.T. Segmentation of the Prostate, Its Zones, Anterior Fibromuscular Stroma, and Urethra on the MRIs and Multimodality Image Fusion Using U-Net Model. *Quant. Imaging Med. Surg.* **2022**, *12*, 4786–4804. [\[CrossRef\]](#)
53. Ekinici, S.; Izci, D.; Eker, E.; Abualigah, L. *An Effective Control Design Approach Based on Novel Enhanced Aquila Optimizer for Automatic Voltage Regulator*; Springer: Berlin/Heidelberg, Germany, 2023; Volume 56.
54. Tsuneki, M. Deep Learning Models in Medical Image Analysis. *J. Oral Biosci.* **2022**, *64*, 312–320. [\[CrossRef\]](#)
55. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *BT—Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
56. Yang, G.; Li, G.; Pan, T.; Kong, Y. Automatic Segmentation of Kidney and Renal Tumor in CT Images Based on 3D Fully Convolutional Neural Network with Pyramid Pooling Module. In *Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR)*, Beijing, Chin, 20–24 August 2018; pp. 3790–3795. [\[CrossRef\]](#)
57. Diniz, J.O.B.; Ferreira, J.L.; Diniz, P.H.B.; Silva, A.C.; de Paiva, A.C. Esophagus Segmentation from Planning CT Images Using an Atlas-Based Deep Learning Approach. *Comput. Methods Programs Biomed.* **2020**, *197*, 105685. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Ducros, N.; Mur, A.L.; Peyrin, F. A Completion Network for Reconstruction from Compressed Acquisition. In *Proceedings of the IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, Iowa City, IA, USA, 3–7 April 2020; pp. 619–623. [\[CrossRef\]](#)
59. Türk, F.; Lüy, M.; Barışçi, N. Kidney and Renal Tumor Segmentation Using a Hybrid V-Net-Based Model. *Mathematics* **2020**, *8*, 1772. [\[CrossRef\]](#)
60. Lin, D.T.; Lei, C.C.; Hung, S.W. Computer-Aided Kidney Segmentation on Abdominal CT Images. *IEEE Trans. Inf. Technol. Biomed.* **2006**, *10*, 59–65. [\[CrossRef\]](#)
61. D’Arco, A.; Ferrara, M.A.; Indolfi, M.; Tufano, V.; Sirleto, L. Implementation of Stimulated Raman Scattering Microscopy for Single Cell Analysis. *Nonlinear Opt. Appl.* **2017**, *10228*, 102280S. [\[CrossRef\]](#)
62. Khalifa, F.; Gimel’farb, G.; Abo El-Ghar, M.; Sokhadze, G.; Manning, S.; McClure, P.; Ouseph, R.; El-Baz, A. A New Deformable Model-Based Segmentation Approach for Accurate Extraction of the Kidney from Abdominal CT Images. In *Proceedings of the 18th IEEE International Conference on Image Processing*, Brussels, Belgium, 11–14 September 2011; pp. 3393–3396. [\[CrossRef\]](#)
63. Yang, G.; Gu, J.; Chen, Y.; Liu, W.; Tang, L.; Shu, H.; Toumoulin, C. Automatic Kidney Segmentation in CT Images Based on Multi-Atlas Image Registration. In *Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Chicago, IL, USA, 26–30 August 2014; pp. 5538–5541. [\[CrossRef\]](#)
64. Cuingnet, R.; Prevost, R.; Lesage, D.; Cohen, L. D.; Mory, B.; Ardon, R. Automatic Detection and Segmentation of Kidneys in 3D CT Images Using Random Forests. *Lect. Notes Comput. Sci.* **2012**, *7512*, 66–74.
65. Jin, C.; Shi, F.; Xiang, D.; Jiang, X.; Zhang, B.; Wang, X.; Zhu, W.; Gao, E.; Chen, X. 3D Fast Automatic Segmentation of Kidney Based on Modified AAM and Random Forest. *IEEE Trans. Med. Imaging* **2016**, *35*, 1395–1407. [\[CrossRef\]](#)
66. Hsiao, C.H.; Lin, P.C.; Chung, L.A.; Lin, F.Y.S.; Yang, F.J.; Yang, S.Y.; Wu, C.H.; Huang, Y.; Sun, T.L. A Deep Learning-Based Precision and Automatic Kidney Segmentation System Using Efficient Feature Pyramid Networks in Computed Tomography Images. *Comput. Methods Programs Biomed.* **2022**, *221*, 106854. [\[CrossRef\]](#) [\[PubMed\]](#)
67. Heller, N.; Sathianathan, N.; Kalapara, A.; Walczak, E.; Moore, K.; Kaluzniak, H.; Rosenberg, J.; Blake, P.; Rengel, Z.; Oestreich, M.; et al. The KiTS19 Challenge Data: 300 Kidney Tumor Cases with Clinical Context, CT Semantic Segmentations, and Surgical Outcomes. *arXiv* **2019**, arXiv:1904.00445.
68. Aresta, G.; Araújo, T.; Kwok, S.; Chennamsetty, S.S.; Safwan, M.; Alex, V.; Marami, B.; Prastawa, M.; Chan, M.; Donovan, M.; et al. BACH: Grand Challenge on Breast Cancer Histology Images. *Med. Image Anal.* **2019**, *56*, 122–139. [\[CrossRef\]](#) [\[PubMed\]](#)
69. Zhao, G.; Ge, W.; Yu, Y. GraphFPN: Graph Feature Pyramid Network for Object Detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, 10–17 October 2021; pp. 2743–2752. [\[CrossRef\]](#)
70. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
71. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. PANet: Path Aggregation Network for Instance Segmentation. *arXiv* **2019**, arXiv:1803.01534v3.
72. Zhang, Z.; Zhang, X.; Peng, C.; Xue, X.; Sun, J. ExFuse: Enhancing Feature Fusion for Semantic Segmentation. *Lect. Notes Comput. Sci.* **2018**, *11214*, 273–288.
73. Lin, D.; Shen, D.; Shen, S.; Ji, Y.; Lischinski, D.; Cohen-Or, D.; Huang, H. ZigzagNet: Fusing Top-down and Bottom-up Context for Object Segmentation. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2019**, *2019*, 7482–7491. [\[CrossRef\]](#)
74. Toshev, A.; Szegedy, C. DeepPose: Human Pose Estimation via Deep Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 23–28 June 2014; pp. 1653–1660.
75. Wu, G.; Ji, X.; Yang, G.; Jia, Y.; Cao, C. Signal-to-Image: Rolling Bearing Fault Diagnosis Using ResNet Family Deep-Learning Models. *Processes* **2023**, *11*, 1527. [\[CrossRef\]](#)
76. Task, C. Se-ResNet with GAN Based Data Augmentation Applied to Acoustic Scene Classification. *Detect. Classif. Acoust. Scenes Events* **2018**, *2018*, 10063424.
77. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [\[CrossRef\]](#)
78. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *Proceedings of the 3rd International Conference on Learning Representations—ICLR 2015*, San Diego, CA, USA, 7–9 May 2015; pp. 1–14.

79. Lecun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
80. Taha, A.A.; Hanbury, A. Metrics for Evaluating 3D Medical Image Segmentation: Analysis, Selection, and Tool. *BMC Med. Imaging* **2015**, *15*, 29. [[CrossRef](#)]
81. Popovic, A.; de la Fuente, M.; Engelhardt, M.; Radermacher, K. Statistical Validation Metric for Accuracy Assessment in Medical Image Segmentation. *Int. J. Comput. Assist. Radiol. Surg.* **2007**, *2*, 169–181. [[CrossRef](#)]
82. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. In Proceedings of the IEEE International Conference on Computer Vision, Honolulu, HI, USA, 21–26 July 2017; pp. 2980–2988.
83. Kodym, O.; Španěl, M.; Herout, A. Segmentation of Head and Neck Organs at Risk Using CNN with Batch Dice Loss. *Lect. Notes Comput. Sci.* **2019**, *11269*, 105–114.
84. Zhang, Y.; Liu, S.; Li, C.; Wang, J. Rethinking the Dice Loss for Deep Learning Lesion Segmentation in Medical Images. *J. Shanghai Jiaotong Univ.* **2021**, *26*, 93–102. [[CrossRef](#)]
85. Milletari, F.; Navab, N.; Ahmadi, S. A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In Proceedings of the 4th International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571. [[CrossRef](#)]
86. Elharrouss, O.; Akbari, Y.; Almaadeed, N.; Al-Maadeed, S. Backbones-Review: Feature Extraction Networks for Deep Learning and Deep Reinforcement Learning Approaches. *arXiv* **2022**, arXiv:2206.08016.
87. Liang, S.; Gu, Y. SRENet: A Spatiotemporal Relationship-Enhanced 2D-CNN-Based Framework for Staging and Segmentation of Kidney Cancer Using CT Images. *Appl. Intell.* **2022**, *53*, 17061–17073. [[CrossRef](#)]
88. Sun, P.; Mo, Z.; Hu, F.; Liu, F.; Mo, T.; Zhang, Y.; Chen, Z. Kidney Tumor Segmentation Based on FR2PAttU-Net Model. *Front. Oncol.* **2022**, *12*, 853281. [[CrossRef](#)] [[PubMed](#)]
89. Abdelrahman, A.; Viriri, S. EfficientNet Family U-Net Models for Deep Learning Semantic Segmentation of Kidney Tumors on CT Images. *Front. Comput. Sci.* **2023**, *5*, 1235622.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

4.2 Conclusion

This paper presents an innovative deep learning model that successfully integrates FPN, SE blocks, and ResNet architecture to improve the diagnosis of kidney tumors from CT images. The model demonstrates enhanced feature extraction and greater diagnostic accuracy by effectively capturing multi-scale image details and focusing on critical regions. The results show that this approach outperforms traditional models, offering significant potential for more precise and reliable kidney tumor detection in clinical settings. The study highlights the model's robustness and sets the groundwork for future improvements in medical image analysis.

Chapter 5

Accurate Kidney Tumor Semantic Segmentation Using Deep Learning Techniques on CT Images

5.1 Introduction

The paper "Accurate Kidney Tumor Semantic Segmentation Using Deep Learning Techniques on CT Images" offers a comprehensive analysis of deep learning methods for segmenting kidney tumors in CT scans. By evaluating three prominent backbones—EfficientNet, SE-ResNet, and DenseNet—across three architectures—U-Net, Feature Pyramid Network (FPN), and LinkNet—the study assesses their effectiveness in kidney tumour segmentation using the KiTS19 dataset

Accurate Kidney Tumor Semantic Segmentation Using Deep Learning Techniques on CT Images

Abubaker Abdelrahman¹, Serestina Viriri^{1*}

¹University of KwaZulu-Natal, South Africa

Submitted to Journal:
Frontiers in Computer Science

Specialty Section:
Computer Vision

Article type:
Original Research Article

Manuscript ID:
1469488

Received on:
02 Aug 2024

Journal website link:
www.frontiersin.org

Review

Scope Statement

In this study, we explore three prominent backbones EfficientNet, SE-ResNet, and DenseNet, with three different architectures U-Net, Feature Pyramid Network (FPN), and LinkNet, for kidney tumor segmentation and conduct a comprehensive performance comparison using the KiTS19 dataset.

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Credit Author Statement

Abubaker Abdelrahman: Conceptualization, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. Serestina Viriri: Conceptualization, Formal Analysis, Investigation, Supervision, Validation, Writing – review & editing.

Keywords

deep learning, Kidney tumor segmentation, SE-ResNet, Densenet, Feature pyramid network, LINKNET

Abstract

Word count: 175

The rapid advancement of artificial intelligence (AI) and its widespread applications in healthcare have brought about a revolutionary transformation in medical image analysis. Among the various AI techniques, Deep Learning (DL) methods, especially convolutional neural networks (CNNs), have exhibited great potential in accurately detecting and characterizing kidney tumors. However, selecting a suitable network architecture and backbone is pivotal in achieving optimal results. In this study, we explore three prominent backbones EfficientNet, SE-ResNet, and DenseNet, with three different architectures U-Net, Feature Pyramid Network (FPN), and LinkNet, for kidney tumor segmentation and conduct a comprehensive performance comparison using the KiTS19 dataset. For the proposed models, performance is evaluated via both Intersection over Union (IoU) and F1-score metrics. The highest segmentation scores were obtained using PFN with (Seresnet50, Efficientnetb4, and Densenet121). Our method verifies that Densenet121 with FPN achieved good results in segmenting kidneys and tumors, with Dice scores of 0.982 and 0.984, respectively. These findings underscore the significance of the backbone choice and demonstrate the effectiveness of the proposed approaches in tackling the challenges of kidney tumor segmentation.

Funding statement

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Ethics statements

Studies involving animal subjects

Generated Statement: No animal studies are presented in this manuscript.

Studies involving human subjects

Generated Statement: No human studies are presented in the manuscript.

Inclusion of identifiable human data

Generated Statement: No potentially identifiable images or data are presented in this study.

Data availability statement

Generated Statement: Publicly available datasets were analyzed in this study. This data can be found here: <https://kits19.grand-challenge.org/data/>.

Accurate Kidney Tumor Semantic Segmentation Using Deep Learning Techniques on CT Images

Abubaker Abdelrahman¹, Serestina Viriri^{1*}

¹School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, 4000, South Africa;

*Corresponding Author: Serestina Viriri viriris@ukzn.ac.za

Abstract

The rapid advancement of artificial intelligence (AI) and its widespread applications in healthcare have brought about a revolutionary transformation in medical image analysis. Among the various AI techniques, Deep Learning (DL) methods, especially convolutional neural networks (CNNs), have exhibited great potential in accurately detecting and characterizing kidney tumors. However, selecting a suitable network architecture and backbone is pivotal in achieving optimal results. In this study, we explore three prominent backbones EfficientNet, SE-ResNet, and DenseNet, with three different architectures U-Net, Feature Pyramid Network (FPN), and LinkNet, for kidney tumor segmentation and conduct a comprehensive performance comparison using the KiTS19 dataset. For the proposed models, performance is evaluated via both Intersection over Union (IoU) and F1-score metrics. The highest segmentation scores were obtained using PFN with (Seresnet50, Efficientnetb4, and Densenet121). Our method verifies that Densenet121 with FPN achieved good results in segmenting kidneys and tumors, with Dice scores of 0.982 and 0.984, respectively. These findings underscore the significance of the backbone choice and demonstrate the effectiveness of the proposed approaches in tackling the challenges of kidney tumor segmentation.

Key terms

Artificial intelligence; deep learning; kidney tumor segmentation; medical imaging; EfficientNet; SE-ResNet; DenseNet; U-net; Feature pyramid network; LinkNet

1. Introduction

Artificial intelligence (AI) is a machine's ability to imitate intelligent human behavior and achieve goals across various environments, enabling intelligent decision-making [1]. The rise of AI has sparked controversy, with IBM CEO Ginni Rometty arguing that AI technologies augment human intelligence, creating a partnership between man and machine that improves human capabilities. AI is a machine's ability to learn from experience, adapt to new inputs, and perform human-like tasks. It was first introduced in the 1950s [2]. AI research focuses on understanding the real world using diverse data types for classification, pattern recognition, and decision-making [3][4]. Processing image/video data aims to detect objects, classify behavior, analyze scenes, and prevent aberrant actions using performance-enhancing features. Techniques include traditional statistical methods, neural networks, and deep learning [5]. Statistical-based methods struggled with pattern identification due to task scenarios, aspects, and data content. Machine Learning (ML) techniques address these challenges, but learning from various features improves performance [6]. ML algorithms fight to process large datasets due to their size, preventing the analysis of specific situations. However, Deep Learning techniques like CNNs have made processing massive datasets feasible. Introducing medically annotated extensive data and high-performance computation have transformed DL in medical imaging. These technologies have helped the health sector analyze and study various body organs, enabling the development of various systems for treating millions of diseases, including kidney disease [7].

Automated computer-aided detection (CAD) and diagnostic systems are crucial for detecting and diagnosing nodules from large volumes of radiological images. These systems assist physicians in identifying malignant or benign nodules and characterizing them. The segmentation task significantly influences outputs, making increasing segmentation accuracy an important goal. These systems help physicians identify and characterize nodules, ensuring accurate and reliable diagnosis [8]. Science and technology can truly be regarded as a boon for human civilization if they contribute to enhancing human life's security, health, livability, and comfort. The worldwide anticipation for powerful CAD systems has been evident, and these systems have become indispensable technologies to address global healthcare needs by aiding in early disease detection and diagnosis [9]. The prevalence of kidney tumors has escalated, making it one of the most widespread health concerns on a global scale. The identification of kidney diseases falls under a multidisciplinary domain that leverages information technology to model human kidney processes effectively and generate dependable diagnostic outcomes [10]. Renal cell carcinoma is the most common genitourinary cancer. Kidney cancer starts from the kidney itself and then spreads slowly or rapidly. Segmenting kidney and tumors allows doctors to make more accurate treatment plans. However, manually segmenting region of interest (ROI) can be time-consuming and

tedious, as radiologists need to tag ROI for each individual slice [11][12]. The progress in kidney tumor imaging and the exponential growth of imaging data sets are propelled by big data storage capabilities. This, in turn, enables early detection and improved care by utilizing ML and DL techniques in studies aimed at enhancing tumor diagnosis and treatment outcomes [13]. Accurate identification of kidneys and tumors from medical images, like CT scans, is crucial for appropriate treatment. Rapid diagnosis and treatment are crucial for cured individuals. Radiologists use images of multiple structures to detect disease and aid in diagnosis. With the growing amount of CT data, researchers are developing reliable automated systems to assist radiologists in diagnosing and planning illnesses [12]. DL task selection for feature extraction can be costly due to specific backbones being used for some tasks, making it difficult to find suitable alternatives [14][15]. This paper describes existing backbones for feature extraction, including DenseNet, EfficientNet, and SE-ResNet. It also discusses their use in computer vision tasks like object detection using architectures like U-Net, FPN, and Linknet. And presents a summary of various deep learning backbones, an overview of computer vision tasks, an evaluation and comparison study, and a summary of the most commonly used backbones for each task.

1.1 Background

1.1.1 Backbone

Object detection stands as a fundamental pillar in the realm of computer vision. Thanks to the rapid advancements in deep convolutional neural networks, object detection performance has witnessed significant improvements [16][17][18][19][20]. One-stage object detectors based on CNNs represent the latest category of object detection models, like YOLO [21][22], SSD [23], and RetinaNet [24], and two-stage detectors, e.g. Faster R-CNN [25], R-FCN [26], FPN [27]. Feature extraction is the primary data analysis step, transforming performance and processing capacity with the machine and DL techniques, including neural networks [4]. Then, convolution neural networks are used for feature extraction in DL models, and selecting the right CNN network is not a random process [5]. The implementation of a DL model depends on the target task and complexity. Proposed networks are used for features extraction and are called backbones. Backbones are recognized architectures or networks trained for various tasks and demonstrate their effectiveness. This section provides a detailed description of each backbone used in DL models.

1.1.2 EfficientNet

Convolutional Neural Networks are designed to fix resources and scale with more to improve accuracy. Resolution refers to image resolution, while depth and width represent layer and channel sizes. Scaling depth can process complex information, but requires gradient

vanishing and resource requirements. Scaling width can process finer information, but can lead to saturation inaccuracy. EfficientNet offers better accuracy and efficiency by scaling resolution, width, and depth uniformly [28]. EfficientNet Networks, a recent architecture, outperform other classification tasks with fewer parameters and FLOPs. They use compound scaling to efficiently scale network width, depth, and resolution. EfficientNet parameters are 8.4x smaller and faster on inference than existing networks. The various versions of EfficientNet can be replaced based on resource capacity and computational cost. EfficientNet-B0 contains 5.3 million parameters, while EfficientNet-B7 has 66 million [3]. EfficientNet is a promising medical image analysis tool. The EfficientNet Family U-Net model presented by Abdelrahman and Viriri. [29] obtained average IoU scores of 0.976 for B0 to 0.980 for B4; while B7 having the highest IoU score for kidney segmentation and B4 is having the highest score for tumors. Akshayet et al. [30] Utilized U-Net architecture and EfficientNet encoder for precise polycystic kidney segmentation on MRI, reducing expert contouring time. Geethanjali et al. [12] The model develops an ideal prediction model for CT images using various encoders. The hybrid Attention-U-Net approach, using , outperforms FPN and LinkNet architectures, with IoU scores of 95.66% for kidneys and 93.86% for tumors. Table 1 shows the outcomes of several Methods while Figure 1 makes a comparison between different segmentation approaches.

Table 1: Shows a summary of results EfficientNet on kidney and kidney tumor.

First Author	Modality	Segmentation Methodology	Kidney IoU Score	Tumor IoU Score	Kidney Dice	Tumor Dice
[12]	CT	Attention U-Net Efficientnetb0	0.951	0.927	-	-
[12]	CT	Attention U-Net Efficientnetb1	0.950	0.928	-	-
[12]	CT	Attention U-Net Efficientnetb2	0.952	0.934	-	-
[12]	CT	Attention U-Net Efficientnetb3	0.951	0.928	-	-

[12]	CT	Attention U-Net Efficientnetb4	0.953	0.933	-	-
[12]	CT	Attention U-Net Efficientnetb5	0.949	0.923	-	-
[12]	CT	Attention U-Net Efficientnetb6	0.953	0.934	-	-
[12]	CT	Attention U-Net Efficientnetb7	0.950	0.925	-	-
[29]	CT	U-Net + Efficientnetb0	0.969	0.960	0.984	0.980
[29]	CT	U-Net + Efficientnetb1	0.963	0.961	0.981	0.980
[29]	CT	U-Net + Efficientnetb2	0.966	0.965	0.983	0.982
[29]	CT	U-Net + Efficientnetb3	0.970	0.960	0.985	0.980
[29]	CT	U-Net + Efficientnetb4	0.974	0.968	0.987	0.984
[29]	CT	U-Net + Efficientnetb5	0.972	0.966	0.986	0.983
[29]	CT	U-Net + Efficientnetb6	0.971	0.963	0.985	0.981
[29]	CT	U-Net + Efficientnetb7	0.977	0.962	0.988	0.981

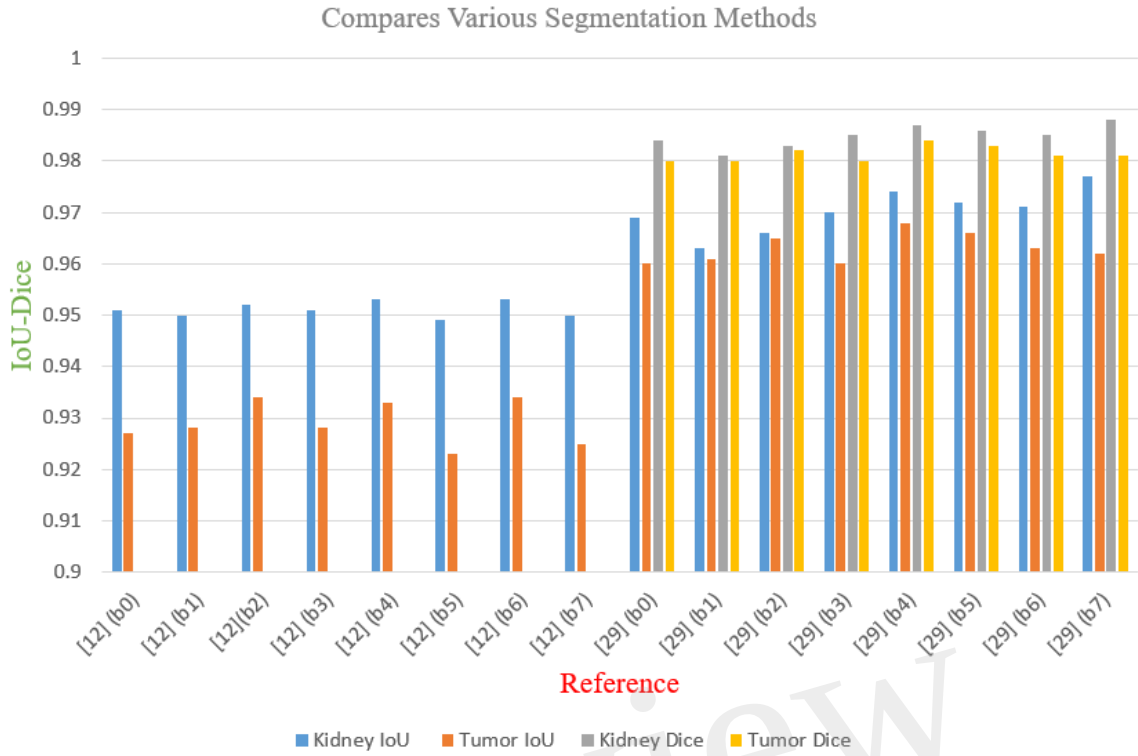


Figure 1: Displays the Outcomes of Several Methods.

1.1.3 SE-ResNet

SE-ResNet [31]. Uses convolution to extract informative features by combining spatial and channel-wise information within local receptive fields, using the SE block [31]. SE-ResNet analyzes interdependencies between convolved feature channels using 1D convolution. It uses a squeeze operation to summarize information and an excitation operation to scale the importance of each mapping. The squeeze operation extracts significant information for each channel, while the excitation operation computes dependencies using a connected layer with a nonlinear function [32]. Limited literature on SE-ResNet's use in kidney and tumor segmentation. Abdelrahman and Viriri. [33] Proposed model uses SE-ResNet to achieve the segmentation of kidney and tumor images; they used FPN architecture, SE-ResNet models that obtained solid IoU scores for background, kidney, and tumor segmentation, with Seresnet50 having most accurate IoU score for kidney segmentation. Table 2 shows a summary of results SE-ResNet on kidney and kidney tumor.

Table 2: Shows a summary of results SE-ResNet on kidney and kidney tumor.

First Author	Modality	Segmentation Methodology	Kidney IoU Score	Tumor IoU Score	Kidney Dice	Tumor Dice
[33]	CT	FPN - SE-ResNet18	0.972	0.971	0.986	0.985
[33]	CT	FPN - SE-ResNet34	0.972	0.971	0.986	0.985
[33]	CT	FPN - SE-ResNet50	0.980	0.984	0.990	0.992
[33]	CT	FPN - SE-ResNet101	0.976	0.982	0.988	0.991
[33]	CT	FPN - SE-ResNet152	0.979	0.984	0.989	0.992

1.1.4 DenseNet

DenseNet is a new convolutional neural network architecture that uses the outputs of previous layers as input for the next layer. This architecture reduces the number of parameters and works with small output channel depths, such as 12 filters per layer. The number of filters used in each convolutional layer is determined by the growth rate of previous layers, making the number of parameters dependent on k . DenseNet has various versions with different number of layers, such as Densenet-121, Densenet-169, Densenet-201, and Densenet-264. The network has an image input size of $224 * 224$ [3]. Gyeongyeon et al. [34] Presented the RBCA module, combining Reverse, Boundary, and Channel Attention, captures various lesion sizes and shapes, achieving high performance compared to other architectures on the KiTS19 dataset, including kidney and tumor. Geethanjali et al. [12] Model develops optimal CT image prediction using encoders and ensemble approach. Table 3 shows a summary of results DenseNet on kidney and kidney tumor.

Table 3: Shows a summary of results DenseNet on kidney and kidney tumor.

First Author	Modality	Segmentation Methodology	Kidney IoU Score	Tumor IoU Score	Kidney Dice	Tumor Dice
[34]	CT	RBCA-NET	-	-	0.941	0.715
[12]	CT	Attention U-Net-Densenet121	0.951	0.930	-	-
[12]	CT	Attention U-Net-Densenet169	0.950	0.927	-	-
[12]	CT	Attention U-Net-Densenet201	0.947	0.924	-	-
[12]	CT	FPN-Densenet121	0.910	0.0	-	-
[12]	CT	FPN-Densenet169	0.807	0.0	-	-
[12]	CT	FPN-Densenet201	0.925	0.0	-	-
[12]	CT	LinkNet-Densenet121	0.903	0.850	-	-
[12]	CT	LinkNet-Densenet169	0.884	0.789	-	-
[12]	CT	LinkNet-Densenet201	0.895	0.820	-	-

1.2 Problem Description

With the limited number of medical experts available [35][36],The lackof available medical experts, the likelihood of inter-observer variability, and the tedious nature of kidney cancer diagnosis highlights the necessity for a capable automated system to accurately classify kidney cancer histopathology images.To have the ability of early detection and better patient outcomes using advanced algorithms and ML techniques to enhance speed and accuracy of cancer diagnosis, safety and effectiveness must be verified before getting into clinical practice. Previous approaches have shown promise [11]. Segmentation methods those rely on the proximity of pixels are known as region-based approaches. In [37] The

authors use the kidney segmentation technique to propose their approach which combines kernel fact hunt and adaptive region-growing to improve attention construction. To prevent nearby organs from influencing deformable models, Erdt et al. [38] use basic constraints. The level-set procedure is employed to create a deformable model in probabilistic form by the authors of [39]; they enabled voxel communication and incorporate a stochastic haste function to enhance segmentation performance. On the other hand, the atlas-based technique uses functional diagrams for consistent annotations and orientation. The paper explores using different network architectures as backbones for DL in kidney tumor analysis, a crucial task in medical image analysis. The choice of architecture depends on factors like problem domain, data availability, computational resources, and desired performance metrics. Researchers can explore different approaches to address weaknesses, such as improving model capture, robustness to noise, and tumor characteristics. This iterative process advances DL and enhances kidney tumor detection and analysis accuracy and reliability.

1.3 Paper Structure

This paper is structured into different sections. Part 2, referred to as "Materials and Methods," covers the dataset used and presents information about the framework adopted for this study. In Part 3, the "Results and Discussion" are presented. Finally, Part 4 delves into the insights obtained from the study and serves as the conclusion of the paper.

2. Material and Method

In this study, segmentation algorithms for identifying and defining kidney tumors are trained and evaluated using a subset of the KiTS19 dataset made up of CT scans of abdomens with kidney tumors. Intersection over Union (IoU) and F1-score measures are used to rate the method's precision and efficacy. The main objective is to accurately and quickly identify and delineate kidney and kidney tumors within CT images. This will help doctors better diagnose and track kidney tumors while also improving patient care and treatment outcomes.

2.1 Dataset

For the study, our data is driven from the KiTS19 dataset [40]. The KiTS19 dataset commonly serves as a comprehensive resource for the segmentation of kidney tumor, providing CT scans with related annotations to train and test machine-learning systems. This dataset is well recognized and benchmark for medical image segmentation models evaluation, specifically for kidney tumors. It includes 210 abdominal CT volumes partitioned into training part and testing part. Every volume has imaging and ground truth labels supplied

in the Neuroimaging Informative Technology Informative (DICOM) format, with all slices orientated axially. This can be found on Kaggle at <https://www.kaggle.com/competitions/body-morphometry-kidney-and-tumor/data> who used the original data obtained from the KiTS19 dataset. The dataset used in this study comprises segments of various sizes, ranging from 29 to 1059 pixels, along with grayscale images measuring 512 by 512 pixels. The specific features of the experimental dataset are available. In this study; a sub dataset including 7,899 PNG images was used. Figure 2 below lists the attributes of the experimental dataset.

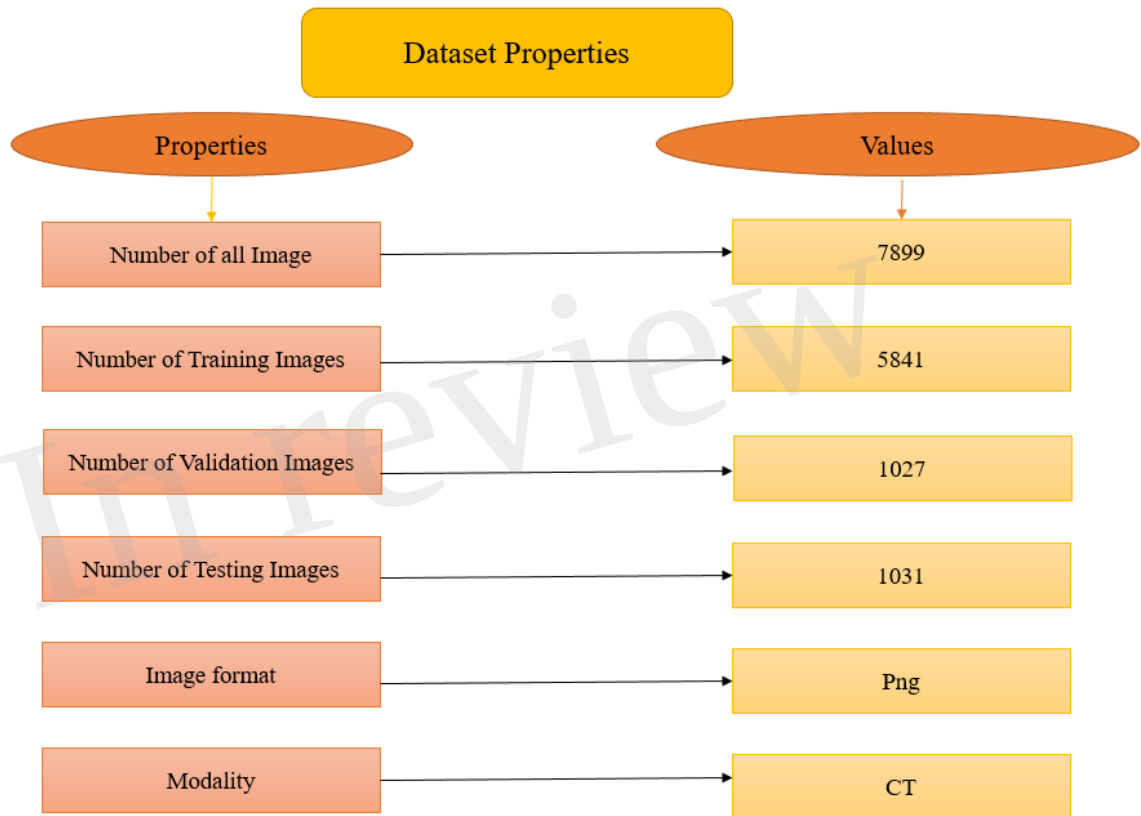


Figure 2: Dataset Properties Used for the Experimentation are cited

2.2 Data Preparation and Pre-Processing

Preprocessing is a vital task for classification CT images, as they are stained with dyes to bring out specific structures and behaviors, It improves the images' regularity and resolution while removing defects and noise that may delay classification. because CNNs are designed to handle smaller inputs, it requires to reduce image resolution to maintain essential details. For preprocessing processes and resizing or reduction of CT images are essential for CNNs to process them effectively [41].in general, CNNs are designed to deal with small inputs. so, there is a need to reduce the resolution of the CT images to a suitable size to work with the network while retaining the essential details. In addition to the steps of preprocessing which

mentioned earlier, an critical step involves resizing or downsampling the CT images to an appropriate dimension for processing with CNNs. the performance of the network can significantly impacted by the size of input images because it requires fixed-size of input. Therefore, downsampling the CT images becomes necessary while maintaining important characteristic and structures within the image. In this paper we start by to converting CT images from DICOM to PNG and set their size to 128×128 as initial processing step.

2.3 U-Net

A well known architecture for image segmentation tasks is U-Net which is a convolutional neural network (CNN) [42]. consists of the contracting path (encoder) and the expansive path (decoder) together draw its U-shape architecture. The encoder side reduces spatial resolution. While catching high-level characteristics and boosting feature channels, the decoder side decreases spatial resolution. The decoder side increases spatial resolution and pixel-wise classification by expecting binary masks for each pixel. U-Net is popular in medical imaging for accurately segmenting organs and structures, as well as satellite image segmentation and semantic segmentation of urban scenes [42]. The U-Net architecture by its encoder-decoder architecture can achieve precise segmentation with fewer training images making it a stable convolutional neural network that used for biomedical image segmentation. The network is made up of 3×3 convolutional layers, having a maximum of 2×2 after every pooling layer and ReLU activation function. Then, convolutional layer 1×1 is attached at the end. The U-Net depends on data augmentation techniques and has the ability to segment a 512×512 image on a modern GPU at significant period of time. It has been applied in medical image segmentation successfully, but 3D convolution is suggested to fully utilize spatial information in 3D images like CT and MRI [42][43].

2.4 Feature Pyramid Network

Feature Pyramid Network predicts the last stratum, unlike U-net [42][44]. FPN is a multiscale feature layer implementation that creates forecast layers at various scales. It includes top-down and bottom-up pathways, as well as side contacts. The bottom-up path encodes hierarchical characteristics, while the final chunk contains the most vigorous characteristics. The top-down algorithm produces high spatial resolution features, while lateral connections enhance feature maps by combining identically sized features from both paths. FPN is a convolutional neural network for semantic segmentation of images [27]. FPN utilizes deep convolutional neural networks' multi-scale hierarchy for high-level semantic feature maps [27]. Experimental segment compares encoder combinations with EfficientNet, SE-ResNets, and DenseNet.

2.5 LinkNet

LinkNet [45]. It is a network that focuses on fast prediction using encoder-decoder structures. It directly propagates spatial information from encoders to decoders, saving time and operations needed for feature restoration. However, LinkNet does not fully utilize longer-range context information, which is crucial for the semantic segmentation of high-resolution aerial imagery. Geethanjali et al. [12] LinkNet uses residuals in the encoder and decoder blocks, partially transferred from the general network. Table 4 shows Encoders utilized in experimentation for feature extraction.

Table 4: Encoders utilized in experimentation for feature extraction.

U-Net	FPN	Link Net
Seresnet50	Efficientnetb4	Densenet121
Densenet121	Densenet121	Densenet169
-	-	Densenet201
-	-	Efficientnetb4
-	-	Seresnet50

2.6 Evaluation Metrics

The study evaluated the segmentation outcomes using two key metrics: F1-score and Intersection over Union (IoU). For each test, these metrics were computed to gauge the effectiveness of the CT image segmentation research. The precision and recall balance are assessed by F1-score, while IoU measures the overlap between the predicted and ground truth regions. By employing these evaluation metrics, the study aimed to gain deeper insights into the performance and accuracy of the segmentation algorithms in CT image analysis. These metrics are expressed as follows:

$$IoU = \frac{TP}{(TP+FP+FN)} \quad [46] \quad (1)$$

$$DSC = \frac{2TP}{2TP + FP + FN} \quad [46] \quad (2)$$

2.7 Proposed solution

The work consists of two parts: the first uses the DenseNet model family as the encoder with Linknet, resulting in superior performance in computer vision tasks. The second part utilizes the best version of EfficientNet in [29], with different architectures, FPN and Linknet, and SE-ResNet in [33] with different architectures, achieving superior results in various tasks.

The proposed method involves two main stages for accurate kidney and tumor segmentation from CT volumes. The first stage involves data preparation, scaling and normalizing CT volumes, and converting them from DICOM images to png images for analysis. The second level includes segmentation via DenseNet with LinkNet model, which effectively extracts features and refines segmentation results using multiscale feature maps. The model simultaneously delineates kidneys and tumor regions, streamlining the process. The optimal segmentation is selected based on intersection-over-union (IoU) scores and F1-score, ensuring accurate delineation of kidney and tumor regions. An open-source resource library, houses cutting-edge for deep-learning models used for image segmentation; including U-Net, PSPNet, FPN, and LinkNet. The Segmentation Models Library provide transfer learning and fine-tuning capabilities on personalized datasets, enabling researchers and practitioners to tackle various image segmentation challenges effectively. The detailed information with source code of these models can be found on GitHub at https://github.com/qubvel/segmentation_models.

2.8 Backbone Architectures

Adequate feature extraction network backbones are critical for the of difficulty and efficiency goal of DL model . More research is required to examine feature extraction networks for DL applications [3]. As an effort to enhance model performance and lower computational expenses, the study evaluates nine feature extraction backbone networks in order to determine which combination works best. Unsuitable backbones can degrade performance and complexity. Figure 3 summarizes the categories Architecture and names of the backbones. DenseNet with Architecture (Linknet, U-net and FPN), EfficientNet with Architecture (Linknet and FPN) and SE-ResNet with Architecture (Linknet and U-net).

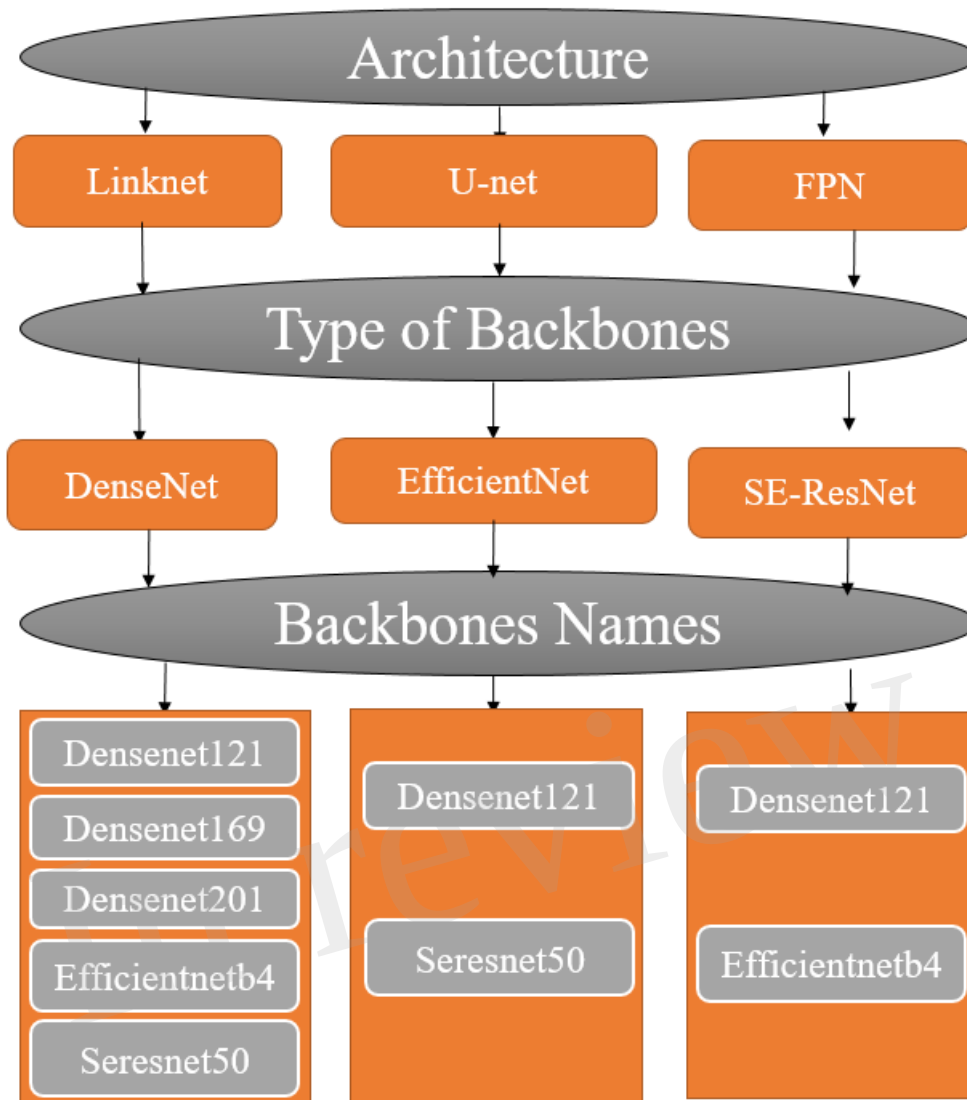


Figure 3: Architecture, Network and Backbones Names.

2.9 Methodology

In this work , a modular design for semantic segmentation of renal tumors has been developed; by combining a semantic segmentation model with thirteen feature extraction layers. The goal is to identify the best segmentation method for kidney tumors. The investigation and evaluation of the best loss function from both model and backbone perspectives, as shown in Figure 4

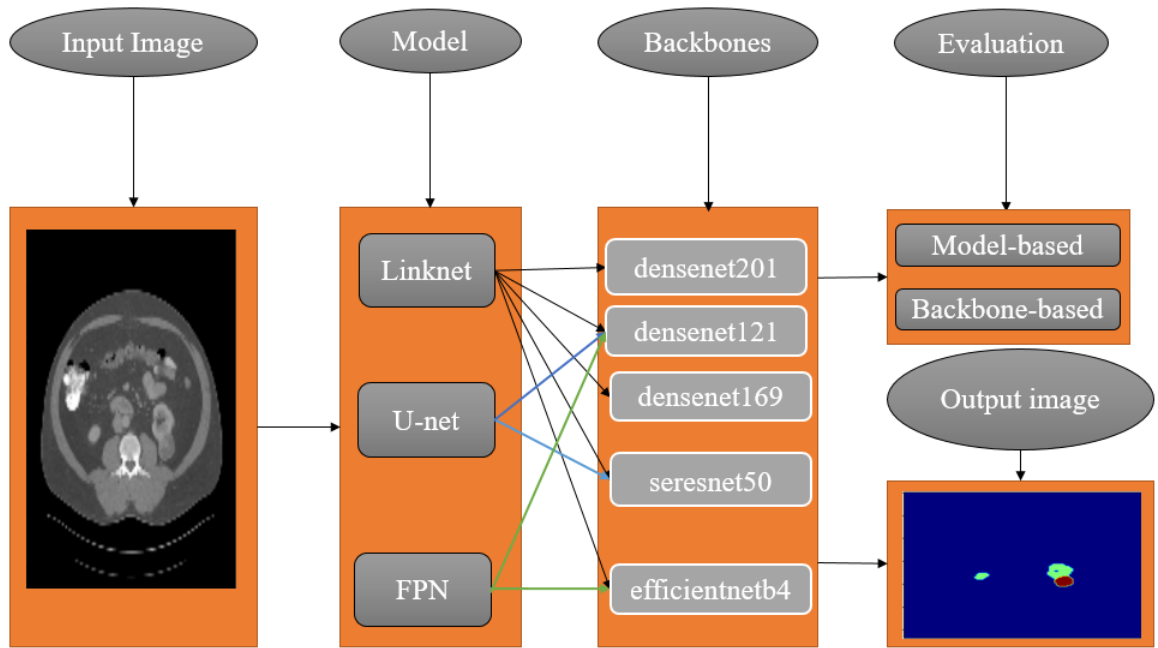


Figure 4: Proposed methodology illustration.

2.10 Model Setup

In the experimentation, every model has been trained for a total of 50 times. The training process halted when the validation loss reached or exceeded 0.0001 to prevent overfitting problem. To speed up convergence, all backbones used weights that had been trained using ImageNet. For a comprehensive overview of the baseline model configurations and hyperparameters of all DL models, refer to Table 5.

Table 5. Models' hyperparameter setup

Hyperparameter	Settings
Activation	'softmax'
Optimizer	Adam
Loss function	Focal Loss, Dice Loss
Learning rate	0.0001
Batch size	32
Epochs	50
Metrics	IoU, F1-score

Input images size 128×128	Input images size 128×128
---------------------------	---------------------------

3. Results and Discussion

Evaluate networks using depth and convolutional filters, DenseNet enhances network performance by increasing depth and examining convolutional filters [47]. EfficientNet balances width, depth, and resolution factors [48], while Seresnet focuses on channel interactions using squeeze and excitation blocks [31]. Both networks improve performance by reusing and strengthening features. The results are divided into two parts: discussing the DenseNet family with LinkNet and evaluating the best version of [29][33] and this paper with different architectures.

3.1 Experimental Results

Table 6 displays, the accuracy of the ensemble model of Densenet121 for Background, Kidney, Kidney Tumor, and Mean IoU outcomes is presented.

Table 6: Ensemble Model outcomes for Densenet121.

IoU Scores	Obtained values
Background IoU Score	0.999
Kidney IOU Score	0.962
Tumor IoU Score	0.954
Mean IoU Score	0.971

3.2 Discussion

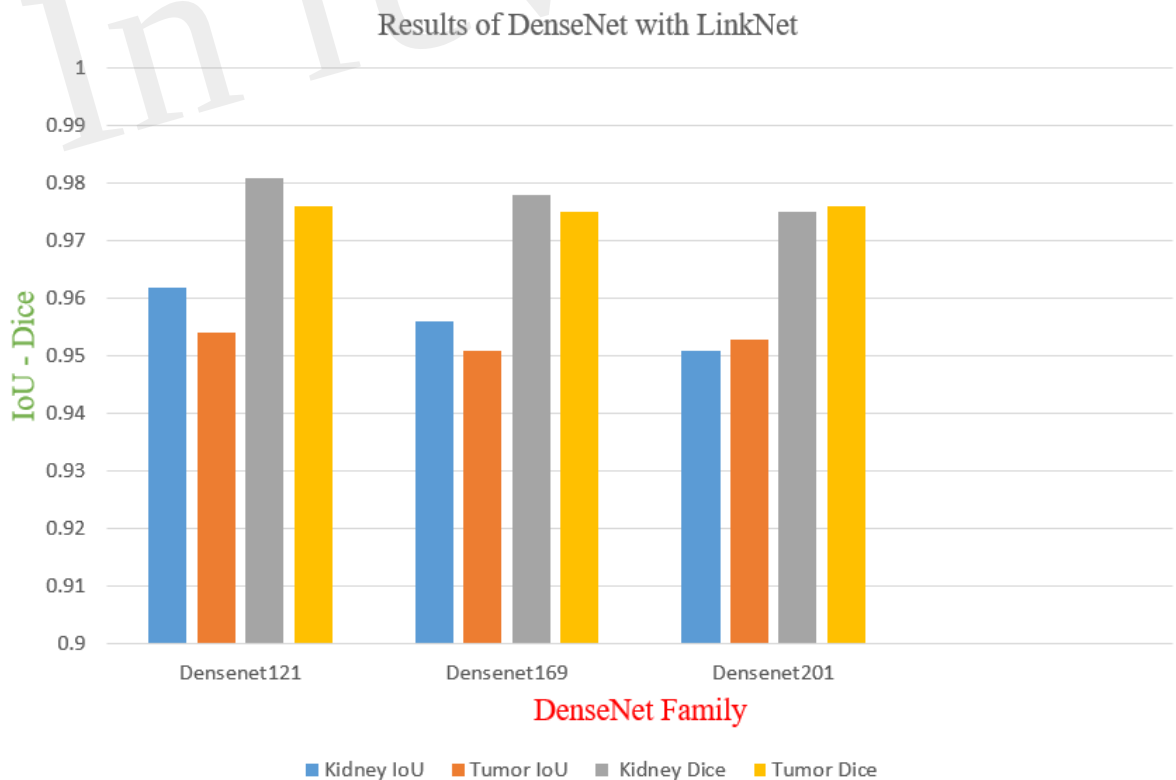
3.2.1 Result Analysis

Using a backbone trained on ImageNet, the research examines the architecture capacity to extract global characteristics from renal cancer CT scans and classify images. as indicated in Table 7.

Table 7: Results for the DenseNet with LinkNet Using Mean IoU and F1-score .

DenseNet	Kidney IoU Score	Tumor IoU Score	Kidney Dice	Tumor Dice
Densenet121	0.962	0.954	0.981	0.976
Densenet169	0.956	0.951	0.978	0.975
Densenet201	0.951	0.953	0.975	0.976

The study compared DenseNet and LinkNet models for segmenting CT images of renal cancer. The models were evaluated using Mean IoU and F1-score to assess their accuracy. The "densenet121" model gained a background IoU score of 0.999, a kidney IoU score of 0.962, and a tumor IoU score of 0.954. Utilizing the F1-score model, the models achieved great IoU scores across several categories; a background Dice score of 0.999, kidney Dice score of 0.981, and tumor Dice score of 0.976. The models "densenet169" and "densenet201" obtained significant IoU and F1-score within the various categories. The findings imply that DenseNet models with LinkNet architecture effectively extract global features from CT images and accurately classify unseen images. as shown in Figure 5.

*Figure 5: DenseNet with LinkNet Results.*

3.2.2 Model Accuracy and Loss

The study evaluated the performance of the models based on two main metrics: Model accuracy and Model loss. Model accuracy indicates how well the model predicts the correct output, while Model loss measures the discrepancy between the predicted output and the actual ground truth values. These metrics provide essential insights into the effectiveness and learning progress of the models during the training and evaluation processes. In Figure 6 and Figure 7, Model accuracy and Model loss Densenet169.

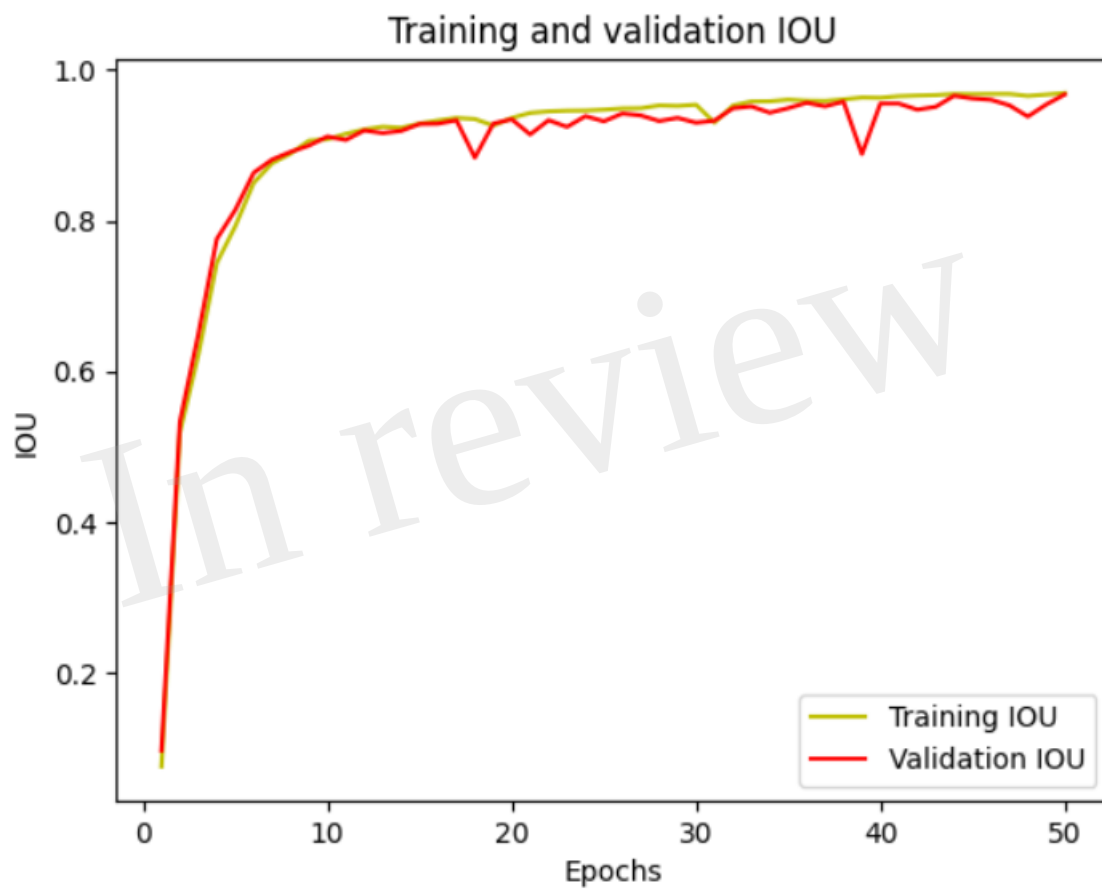


Figure 6. Model IoU for Densenet169

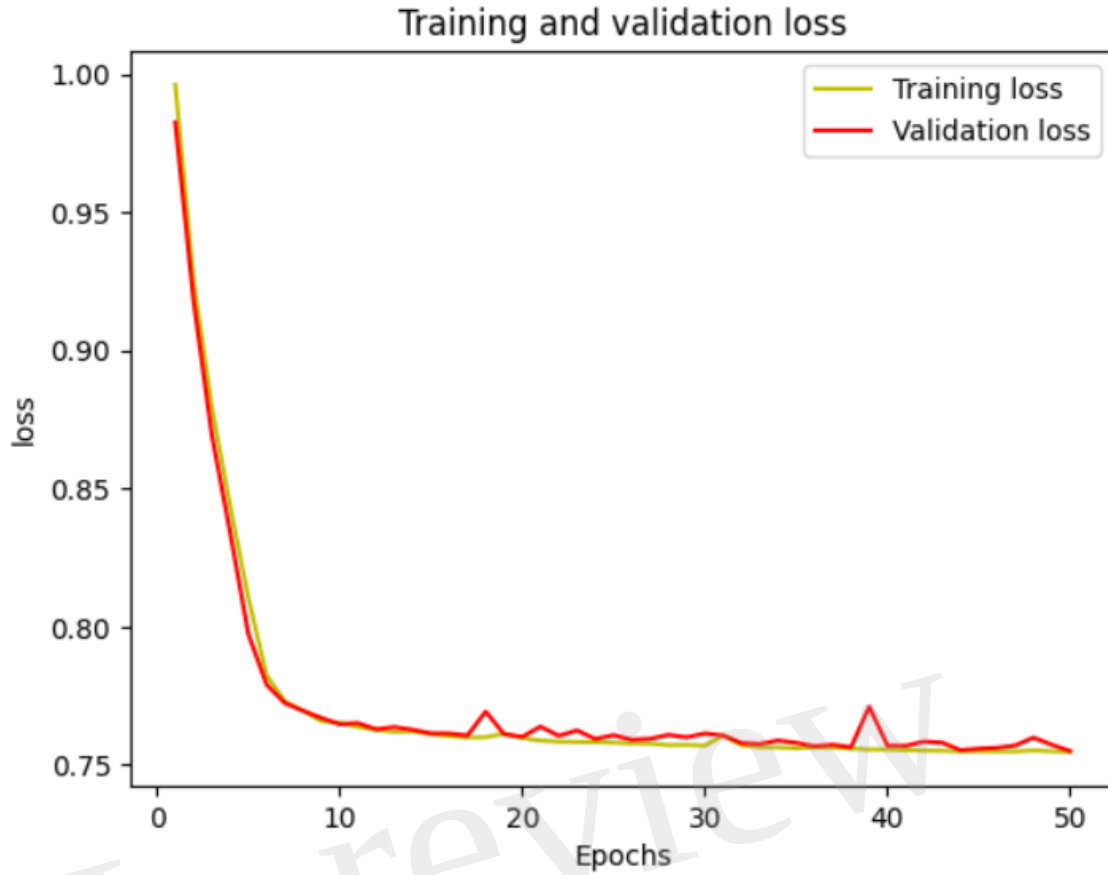


Figure 7. Model Loss for Densenet169

Here, we describe the performance evaluation of loss and IoU score using the DenseNet model. Loss assesses how successfully the model shrink the gap between the expected and actual values, whereas the IoU score assesses the overlap between predicted and ground truth bounding boxes. DenseNet121, densenet169, and densenet201 show consistent performance in terms of loss, with lower values indicating better prediction accuracy. However, IoU score variation exists, with Densenet121 achieving the highest score of 0.969, making it the most accurate model for object detection. It is crucial to consider factors like model size and computational complexity when selecting the most appropriate model for a specific task.

3.2.3 Prediction On Validation Images

An instance of kidney and kidney tumor segmentation using the Densenet121 system is shown in Figure 8. The annotation images show the ground truth descriptions of the kidney and tumor regions, highlighted in green and blue. These annotations are used to evaluate the segmentation model's effectiveness. On the other hand, the result images show the model's kidney and tumor segmentation estimations. The estimated kidney region is marked in green,

and the tumor region is shown in brown. These images visually reflect the segmentation results of the Densenet121 model, illustrating its ability to accurately identify and delineate kidney and tumor regions within CT scans. Color-coded visualizations make evaluating the model's efficacy in kidney and kidney tumor segmentation easier.

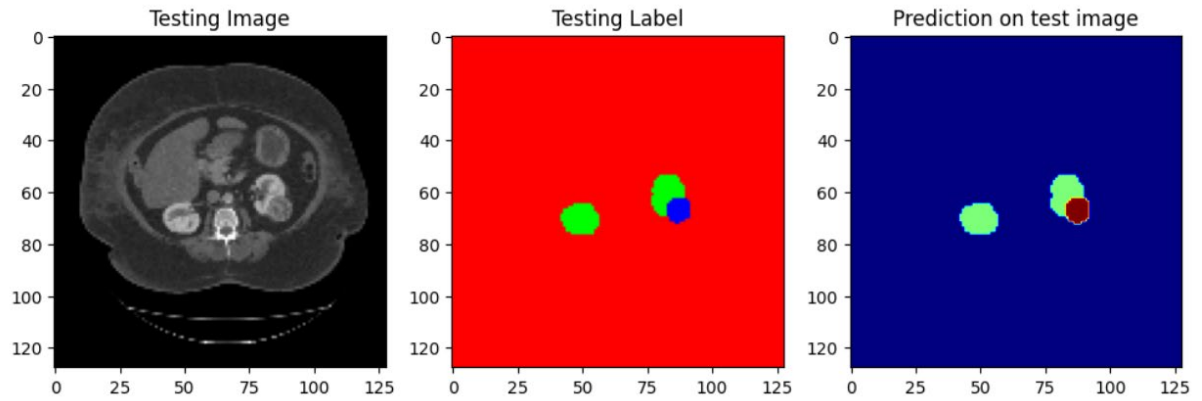


Figure 8. The example of kidney and tumor segmentation

Table 8. Evaluating Pretrained Proposed Modles with Prior Methods

Reference	Architecture	Kidney Dice	Tumor Dice	Kidney IoU Scores	Tumor IoU Scores
[34]	RBCA-NET	0.941	0.715	-	-
[12]	Attention U-Net-Densenet121	-	-	0.951	0.930
[12]	Attention U-Net-Densenet169	-	-	0.950	0.927
[12]	Attention U-Net-Densenet201	-	-	0.947	0.924
[12]	FPN-Densenet121	-	-	0.910	0.0
[12]	FPN-Densenet169	-	-	0.807	0.0
[12]	FPN-Densenet201	-	-	0.925	0.0

[12]	LinkNet-Densenet121	-	-	0.903	0.850
[12]	LinkNet-Densenet169	-	-	0.884	0.789
[12]	LinkNet-Densenet201	-	-	0.895	0.820
Proposed	LinkNet-Densenet121	0.981	0.976	0.962	0.954
Proposed	LinkNet-Densenet169	0.978	0.975	0.956	0.951
Proposed	LinkNet-Densenet201	0.975	0.976	0.951	0.953

Based on the results shown in Table 8, our approach utilizing the LinkNet with different versions of the DenseNet model (Densenet121, Densenet169, and Densenet201) achieves high kidney and tumor detection scores. With a kidney score of 0.962 and a tumor score of 0.954 utilizing IoU, LinkNet-Densenet121 performs the best out of all the designs. This suggests that our LinkNet-Densenet121-based method for kidney and tumor detection is very precise and efficient. The scores given in the table provide a comparative examination of the various methods employed in kidney and tumor identification, thus it's vital to keep in mind that different architectures perform differently throughout the tasks. Figure 9 illustrates the comparison of different segmentation approaches visually.

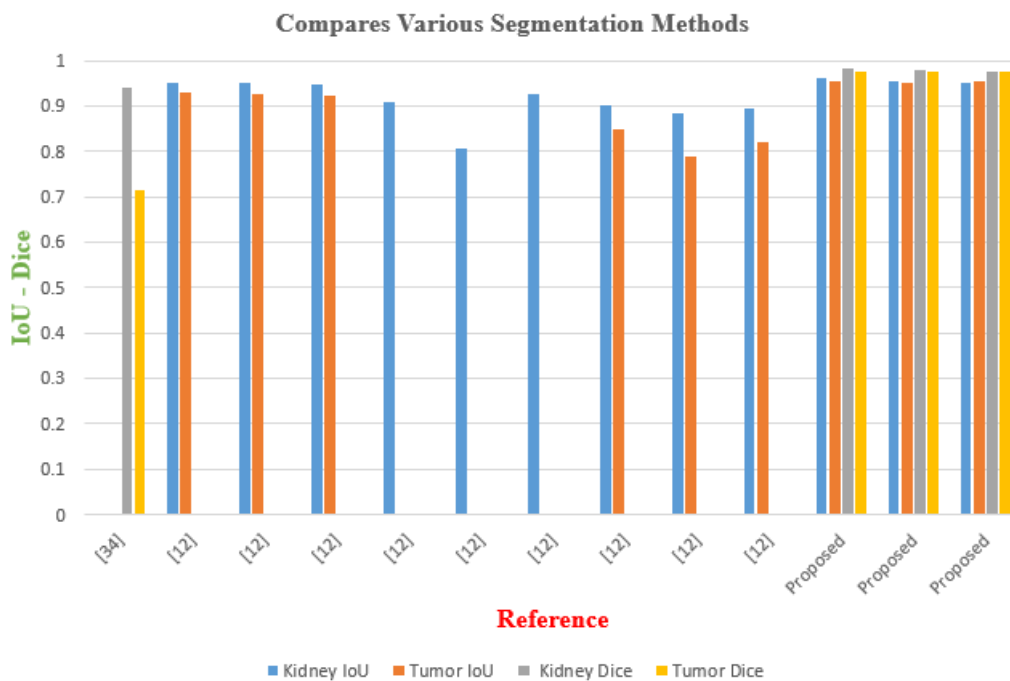


Figure 9: Comparison of Different Methods for Kidney Tumor

3.2.4 Comparing Kidney and Kidney Tumor Models Utilizing Various Backbones

An In-Depth Analysis of Kidney and Kidney Tumor Models: A Comparative Study of EfficientNet and SE-ResNet Based on Two Published Papers, with the Identification of the Optimal DenseNet Version in this Research. We undertake a thorough evaluation to identify the most effective models for each architecture. To achieve this, we utilize three different backbones trained on the ImageNet dataset in conjunction with three distinct architectures. The primary concern of the investigation centers on the capability of the built architecture to extract global characteristics from CT scans of kidney cancer and accurately classify those images. A comparative analysis using multiple models ultimately determines which model architecture yields the best kidney and kidney tumor analysis results. The findings from this study have the potential to contribute significantly to the field of medical image analysis and improve diagnostic accuracy and treatment planning for renal cancer patients.

Table 9: Results for the three different backbones with three different Architectures Using Mean IoU and F1-score.

Reference	Architecture	Backbone	Kidney IoU-score	Tumor IoU-score	Kidney Dice	Tumor Dice
[33]	FPN	Seresnet50	0.980	0.984	0.990	0.992
Proposed	U-Net	Seresnet50	0.977	0.971	0.988	0.985
Proposed	LinkNet	Seresnet50	0.968	0.971	0.984	0.985
[29]	U-Net	Efficientnetb4	0.974	0.968	0.987	0.984
Proposed	FPN	Efficientnetb4	0.980	0.974	0.990	0.987
Proposed	LinkNet	Efficientnetb4	0.960	0.955	0.980	0.977
Proposed	LinkNet	Densenet121	0.962	0.954	0.981	0.976
Proposed	U-Net	Densenet121	0.937	0.948	0.967	0.973
Proposed	FPN	Densenet121	0.965	0.969	0.982	0.984

In Table 9, we present an extensive evaluation of kidney and kidney tumor analysis, wherein we compare the performance of various architectures and their corresponding backbones. Two fundamental metrics, Intersection over Union (IoU) and F1-scores, are employed to assess the effectiveness of each architecture-backbone combination in addressing the challenges inherent in kidney and kidney tumor segmentation. The IoU scores provide valuable insights into the accuracy of segmentation, while the F1-scores offer crucial

information regarding classification performance. Utilizing these metrics gives us a comprehensive understanding of how well each model performs in segmenting kidney and kidney tumor regions. The table reveals that the models exhibit differing performance levels based on the specific combination of architecture and backbone used for kidney and kidney tumor segmentation. The FPN with Seresnet50 backbone, as mentioned in Reference [33], demonstrates the best results, achieving the highest IoU and Dice scores for kidney and tumor segmentation.

Nevertheless, some of the proposed models also exhibit competitive performance. Notably, combining Efficientnetb4 with FPN yields improved results compared to the findings mentioned in Reference [29].

Similarly, the combination of Densenet121 with FPN outperforms Densenet121 with LinkNet. These findings provide valuable insights for researchers and practitioners in selecting the most suitable architecture and backbone combination for accurate and effective kidney and kidney tumor segmentation. Further investigation and comparison will continue to drive advancements in medical image analysis, benefiting patient diagnosis and treatment planning in kidney-related diseases.

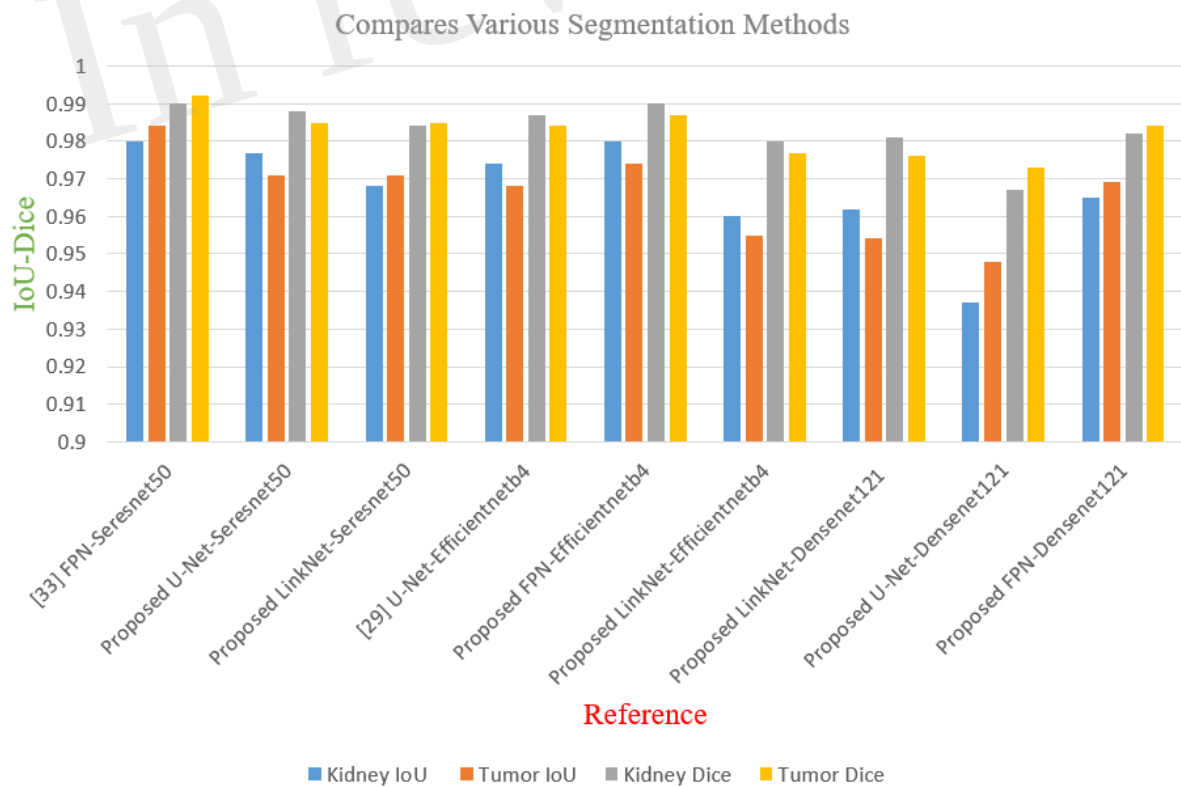


Figure 10: Comparison of Different Methods for Kidney Tumor

4. Conclusion

In conclusion, the research describes CT slice segmentation techniques for tumor and kidney segmentation that employ three backbones and architectures. The suggested method includes a data preparation, segmentation, and ensemble stage for CT image classification, as well as preprocessing, downsampling, resizing, and evaluating image segmentation with IoU and F1-score measures.

This study evaluated the performance of many DL designs for segmenting CT images of renal carcinoma, including DenseNet with LinkNet, Efficientnetb4, and Seresnet50. The results demonstrated how successfully and precisely the DenseNet models with LinkNet architecture extracted and classified global information from the images. The DenseNet models achieved high IoU and F1 scores, indicating successful image segmentation and classification in the images. "densenet121" regularly outperformed the other DenseNet models, achieving the highest F1 and IoU scores.

The study also compared architectures with different backbones, including FPN, U-net, and LinkNet, using mean IoU and F1-score metrics. The ensemble approach with FPN architecture and Seresnet50 backbone yielded the best results regarding IoU scores, indicating accurate segmentation of kidneys and tumors. The Efficientnetb4 backbone with FPN architecture also performed well, outperforming other combinations in some cases. Our method confirmed that Densenet121 with FPN got good results in kidneys and tumor.

Overall, this research provides valuable insights into the performance of deep-learning models for segmenting renal cancer CT images. It helps model designers evaluate the performance of their novel architectures and select appropriate combinations of backbones and architectures for specific tasks. The findings contribute to advancements in the field of medical image analysis and have implications for future research and practical applications in image classification and object detection.

- [1] P. Aghion, B.F. Jones, C.I. Jones, Artificial Intelligence and Economic Growth, 2019.
- [2] Y. Duan, J.S. Edwards, Y.K. Dwivedi, Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda, Int. J. Inf. Manage. 48 (2019) 63–71, <https://doi.org/10.1016/j.ijinfomgt.2019.01.021>.
- [3] O. Elharrouss, Y. Akbari, N. Almaadeed, S. Al-Maadeed, Backbones-Review: Feature Extraction Networks for Deep Learning and Deep Reinforcement Learning Approaches, (2022). <http://arxiv.org/abs/2206.08016>.
- [4] M. Pietikäinen, O. Silven, Challenges of Artificial Intelligence -- From Machine Learning and Computer Vision to Emotional Intelligence, 2022.

<http://arxiv.org/abs/2201.01466>.

- [5] J. Zhou, L. Liu, W. Wei, J. Fan, Network Representation Learning: From Preprocessing, Feature Extraction to Node Embedding, *ACM Comput. Surv.* 55 (2023), <https://doi.org/10.1145/3491206>.
- [6] O. Elharrouss, A. Abbad, D. Moujahid, J. Riffi, H. Tairi, A block-based background model for moving object detection, *Electron. Lett. Comput. Vis. Image Anal.* 15 (2016) 17–31, <https://doi.org/10.5565/REV/ELCVIA.855>.
- [7] S. Shubham, N. Jain, V. Gupta, S. Mohan, M.M. Ariffin, A. Ahmadian, Identify glomeruli in human kidney tissue images using a deep learning approach, *Soft Comput.* 27 (2023) 2705–2716, <https://doi.org/10.1007/s00500-021-06143-z>.
- [8] R.J. Suji, W.W. Godfrey, J. Dhar, Comparing different deep learning backbones for segmentation of lung nodules, 2021 5th Conf. Inf. Commun. Technol. CICT 2021 (2021) 1–5, <https://doi.org/10.1109/CICT53865.2020.9672439>.
- [9] R.F. Mansour, Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy, *Biomed. Eng. Lett.* 8 (2018) 41–57, <https://doi.org/10.1007/s13534-017-0047-y>.
- [10] M. Gharaibeh, D. Alzu'bi, M. Abdullah, I. Hmeidi, M.R. Al Nasar, L. Abualigah, A.H. Gandomi, Radiology Imaging Scans for Early Diagnosis of Kidney Tumors: A Review of Data Analytics-Based Machine Learning and Deep Learning Approaches, *Big Data Cogn. Comput.* 6 (2022) 1–29, <https://doi.org/10.3390/bdcc6010029>.
- [11] A. Abdelrahman, S. Viriri, Kidney Tumor Semantic Segmentation Using Deep Learning: A Survey of State-of-the-Art, *J. Imaging* 8 (2022), <https://doi.org/10.3390/jimaging8030055>.
- [12] T.M. Geethanjali, Minavathi, M.S. Dinesh, Ensemble Approach for Improving Kidney Tumors Segmentation Performance on Ct Images Using Deep Learning Models, *Indian J. Comput. Sci. Eng.* 13 (2022) 467–476, <https://doi.org/10.21817/indjcse/2022/v13i2/221302130>.
- [13] A. Myronenko, A. Hatamizadeh, 3D Kidneys and Kidney Tumor Semantic Segmentation using Boundary-Aware Networks, (2019) 1–8. <http://arxiv.org/abs/1909.06684>.
- [14] S.S.A. Zaidi, M.S. Ansari, A. Aslam, N. Kanwal, M. Asghar, B. Lee, A survey of modern deep learning based object detection models, *Digit. Signal Process. A Rev. J.* 126 (2022) 103514, <https://doi.org/10.1016/j.dsp.2022.103514>.
- [15] K. Tong, Y. Wu, Deep learning-based detection from the perspective of small or tiny objects: A survey, *Image Vis. Comput.* 123 (2022) 104471, <https://doi.org/10.1016/j.imavis.2022.104471>.
- [16] Z. Li, C. Peng, G. Yu, X. Zhang, Y. Deng, J. Sun, DetNet: A Backbone network for Object Detection, (2018) 1–17. <http://arxiv.org/abs/1804.06215>.

- [17] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, *Commun. ACM* 60 (2017) 84–90, <https://doi.org/10.1145/3065386>.
- [18] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.* (2015) 1–14.
- [19] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015: pp. 1–9.
- [20] S. Xie, R. Girshick, P. Doll, Aggregated Residual Transformations for Deep Neural Networks <http://arxiv.org/abs/1611.05431v2>, *Conf. Comput. Vis. Pattern Recognit.* (2017) 1492–1500.
- [21] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016: pp. 779–788.
- [22] J. Redmon, A. Farhadi, YOLO9000: better, faster, stronger, in: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017: pp. 7263–7271.
- [23] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, Ssd: Single shot multibox detector, in: *Comput. Vision—ECCV 2016 14th Eur. Conf. Amsterdam, Netherlands, Oct. 11–14, 2016, Proceedings, Part I 14*, Springer, 2016: pp. 21–37.
- [24] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: *Proc. IEEE Int. Conf. Comput. Vis.*, 2017: pp. 2980–2988.
- [25] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, *Adv. Neural Inf. Process. Syst.* 28 (2015).
- [26] J. Dai, Y. Li, K. He, J. Sun, R-fcn: Object detection via region-based fully convolutional networks, *Adv. Neural Inf. Process. Syst.* 29 (2016).
- [27] T.Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017-Janua* (2017) 936–944, <https://doi.org/10.1109/CVPR.2017.106>.
- [28] M.A. Haq, I. Khan, A. Ahmed, S.M. Eldin, A. Alshehri, N.A. Ghamry, DCNNBT: a novel deep convolution neural network-based brain tumor classification model, *Fractals* 31 (2023), <https://doi.org/10.1142/s0218348x23401023>.
- [29] A. Sharbatdaran, D. Romano, K. Teichman, H. Dev, S.I. Raza, A. Goel, M.C. Moghadam, J.D. Blumenfeld, J.M. Chevalier, D. Shimonov, G. Shih, Y. Wang, M.R. Prince, Deep Learning Automation of Kidney, Liver, and Spleen Segmentation for Organ Volume Measurements in Autosomal Dominant Polycystic Kidney Disease, *Tomography* 8 (2022) 1804–1819, <https://doi.org/10.3390/tomography8040152>.

- [30] A. Goel, G. Shih, S. Riyahi, S. Jeph, H. Dev, R. Hu, D. Romano, K. Teichman, J.D. Blumenfeld, I. Barash, I. Chicos, H. Rennert, M.R. Prince, Deployed Deep Learning Kidney Segmentation for Polycystic Kidney Disease MRI, *Radiol. Artif. Intell.* 4 (2022) 1–7, <https://doi.org/10.1148/RYAI.210205>.
- [31] J. Hu, Squeeze-and-Excitation_Networks_CVPR_2018_paper.pdf, *Cvpr* (2018) 7132–7141. http://openaccess.thecvf.com/content_cvpr_2018/html/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.html.
- [32] T. Chen, H. Qin, X. Li, W. Wan, W. Yan, A Non-Intrusive Load Monitoring Method Based on Feature Fusion and SE-ResNet, *Electron.* 12 (2023), <https://doi.org/10.3390/electronics12081909>.
- [33] Z. Zhu, S. Lu, S.H. Wang, J.M. Górriz, Y.D. Zhang, BCNet: A Novel Network for Blood Cell Classification, *Front. Cell Dev. Biol.* 9 (2022) 1–12, <https://doi.org/10.3389/fcell.2021.813996>.
- [34] G. Hwang, H. Yoon, Y. Ji, S.J. Lee, RBCA-Net: Reverse Boundary Channel Attention Network for Kidney Tumor Segmentation in CT images, *Int. Conf. ICT Converg.* 2022-Octob (2022) 2114–2117, <https://doi.org/10.1109/ICTC55196.2022.9952992>.
- [35] I. Kandel, M. Castelli, How deeply to fine-tune a convolutional neural network: a case study using a histopathology dataset, *Appl. Sci.* 10 (2020) 3359.
- [36] C. Munien, S. Viriri, Classification of Hematoxylin and Eosin-Stained Breast Cancer Histology Microscopy Images Using Transfer Learning with EfficientNets, *Comput. Intell. Neurosci.* 2021 (2021), <https://doi.org/10.1155/2021/5580914>.
- [37] D.T. Lin, C.C. Lei, S.W. Hung, Computer-aided kidney segmentation on abdominal CT images, *IEEE Trans. Inf. Technol. Biomed.* 10 (2006) 59–65, <https://doi.org/10.1109/TITB.2005.855561>.
- [38] A. D’Arco, M.A. Ferrara, M. Indolfi, V. Tufano, L. Sirleto, Implementation of stimulated Raman scattering microscopy for single cell analysis, *Nonlinear Opt. Appl.* X 10228 (2017) 102280S, <https://doi.org/10.1117/12.2266776>.
- [39] F. Khalifa, G. Gimel’farb, M. Abo El-Ghar, G. Sokhadze, S. Manning, P. McClure, R. Ouseph, A. El-Baz, A new deformable model-based segmentation approach for accurate extraction of the kidney from abdominal CT images, *Proc. - Int. Conf. Image Process. ICIP* (2011) 3393–3396, <https://doi.org/10.1109/ICIP.2011.6116438>.
- [40] N. Heller, N. Sathianathen, A. Kalapara, E. Walczak, K. Moore, H. Kaluzniak, J. Rosenberg, P. Blake, Z. Rengel, M. Oestreich, J. Dean, M. Tradewell, A. Shah, R. Tejapaul, Z. Edgerton, M. Peterson, S. Raza, S. Regmi, N. Papanikolopoulos, C. Weight, The KiTS19 Challenge Data: 300 Kidney Tumor Cases with Clinical Context, CT Semantic Segmentations, and Surgical Outcomes, (2019) 1–14. <http://arxiv.org/abs/1904.00445>.
- [41] G. Aresta, T. Araújo, S. Kwok, S.S. Chennamsetty, M. Safwan, V. Alex, B. Marami, M. Prastawa, M. Chan, M. Donovan, G. Fernandez, J. Zeineh, M. Kohl, C. Walz, F. Ludwig, S. Braunewell, M. Baust, Q.D. Vu, M.N.N. To, E. Kim, J.T. Kwak, S. Galal, V. Sanchez-

- Freire, N. Brancati, M. Frucci, D. Riccio, Y. Wang, L. Sun, K. Ma, J. Fang, I. Kone, L. Boulmane, A. Campilho, C. Eloy, A. Polónia, P. Aguiar, BACH: Grand challenge on breast cancer histology images, *Med. Image Anal.* 56 (2019) 122–139, <https://doi.org/10.1016/j.media.2019.05.010>.
- [42] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation BT - *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, in: N. Navab, J. Hornegger, W.M. Wells, A.F. Frangi (Eds.), Springer International Publishing, Cham, 2015: pp. 234–241.
- [43] D. Štern, H. Bischof, M. Urschler, C. Payer, D. Darko Štern, Regressing Heatmaps for Multiple Landmark Localization using CNNs Automatic age estimation from skeletal and dental MRI data using machine learning View project SEE PROFILE Regressing Heatmaps for Multiple Landmark Localization using CNNs, *Miccai* (2016). <https://www.researchgate.net/publication/304024957>.
- [44] P. Isola, J. Xiao, A. Torralba, A. Oliva, What makes an image memorable?, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* (2011) 145–152, <https://doi.org/10.1109/CVPR.2011.5995721>.
- [45] A. Chaurasia, E. Culurciello, LinkNet: Exploiting encoder representations for efficient semantic segmentation, 2017 IEEE Vis. Commun. Image Process. VCIP 2017 2018-Janua (2018) 1–4, <https://doi.org/10.1109/VCIP.2017.8305148>.
- [46] S. Pandey, P.R. Singh, J. Tian, Biomedical Signal Processing and Control An image augmentation approach using two-stage generative adversarial network for nuclei image segmentation, *Biomed. Signal Process. Control* 57 (2020) 101782, <https://doi.org/10.1016/j.bspc.2019.101782>.
- [47] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017: pp. 4700–4708.
- [48] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: *Int. Conf. Mach. Learn.*, PMLR, 2019: pp. 6105–6114.

5.2 Conclusion

This paper, "Accurate Kidney Tumor Semantic Segmentation Using Deep Learning Techniques on CT Images," published in *Frontiers in Computer Science*, presents a significant advancement in the field of medical image analysis. The study meticulously compares three deep learning backbones—EfficientNet, SE-ResNet, and DenseNet—across three different architectures—U-Net, Feature Pyramid Network (FPN), and LinkNet—using the KiTS19 dataset to enhance kidney tumor segmentation accuracy. The findings reveal that DenseNet121 integrated with FPN consistently delivers the highest performance, achieving superior Dice scores of 0.982 for kidneys and 0.984 for tumors. This demonstrates its robustness in accurately segmenting both kidneys and tumors. The research underscores the critical role of selecting the optimal backbone and architecture to achieve the best results in semantic segmentation tasks. Overall, the paper provides valuable insights into the effectiveness of various deep-learning models for kidney tumour segmentation, offering guidance for future research and practical applications. The demonstrated success of DenseNet121 with FPN highlights its potential for improving diagnostic accuracy and efficiency in clinical settings, advancing the field of medical image analysis.

Chapter 6

Results and Discussion

6.1 Introduction

This chapter provides a comprehensive discussion of the overall findings presented throughout this thesis. It synthesizes the results obtained across all the studies included in this work to illustrate how the challenges associated with [specific topic/area] have been addressed. A summary of the key challenges identified in the introduction and literature review is provided, followed by detailed sections outlining the approaches undertaken to overcome these challenges.

6.2 Summary of the Overall Results

Chapters 1 and 2 presented the introduction and literature review, respectively. These chapters' various techniques and computer-aided diagnosis (CAD) systems for analyzing kidney tumour CT images were reviewed, highlighting their limitations. Chapter 2 included the paper titled "Kidney Tumor Semantic Segmentation Using Deep Learning: A Survey of State-of-the-Art." This paper critically examined state-of-the-art methods for analyzing kidney tumor images to detect and diagnose various kidney tumor diseases, comparing these modern approaches with traditional techniques. Despite advancements, several challenges affecting the performance of these methods in analyzing kidney tumor images for accurate detection and diagnosis were identified. The findings from Chapters 1 and 2, which underscore the necessity of this research, are summarized below:

1. Heavy Reliance on Parameter Tuning in Kidney tumor Detection

Deep learning models for kidney tumor detection rely heavily on the tuning of millions of parameters to achieve high performance. This process can lead to overfitting, poor generalization, and significant computational resource demands. The need for extensive memory and processing power to classify kidney tumors accurately can make these systems impractical for real-time use or deployment in resource-limited environments. To improve efficiency and diagnostic accuracy, especially for life-saving applications, there is a growing need for lightweight frameworks that reduce computational and storage complexities while maintaining high classification performance for kidney tumor detection.

2. Automatic and Quick Detection of Kidney Tumors

Timely detection of kidney tumors is critical for improving patient outcomes, as

early diagnosis significantly reduces the risk of progression to advanced stages. kidney tumors, especially malignant ones, can be life-threatening if not detected early. Therefore, developing a robust and efficient framework for the automatic detection of kidney tumors is essential. Such a system would need to accurately classify tumor types—benign or malignant—quickly and reliably, enabling early intervention and potentially saving lives. This emphasizes the need for advanced deep learning models capable of swift and precise tumor identification from imaging data like CT scans.

3. Variability in Tumor Characteristics

Kidney tumors exhibit a wide range of appearances in CT scans, varying in size, shape, density, and location. This variability complicates segmentation and detection, as deep learning models must be robust enough to handle these differences. Some tumors may be small or located in hard-to-reach regions, while others may have irregular shapes or boundaries that are not easily distinguishable from surrounding tissue.

By addressing these challenges, the development of more accurate, efficient, and deployable deep learning models for kidney tumor detection and segmentation can significantly improve clinical outcomes and patient care.

6.2.1 Investigating Lightweight EfficientNet Models for Precise Semantic Segmentation of CT Images

Chapter three addresses the challenges of diagnosing kidney cancer, such as a shortage of medical experts, variability in expert opinions, and the time-intensive nature of manual diagnosis. It emphasizes the need for automated systems that leverage advanced machine learning techniques to improve diagnostic speed and accuracy, ultimately enhancing early detection and patient outcomes. The study investigates the performance of eight lightweight EfficientNet architectures, aiming to optimize resource usage while maintaining high accuracy. These models show promise in achieving results comparable to state-of-the-art methods, with EfficientNet-B4 excelling in tumor detection and EfficientNet-B7 in kidney detection, achieving a mean IoU accuracy of 0.980. The research highlights EfficientNets as effective tools for kidney cancer diagnosis, offering simplicity, faster training times, and reliable performance.

The results from this study demonstrate the effectiveness of integrating EfficientNet family architectures into U-Net models for kidney tumor segmentation in CT images. The models showcased strong segmentation performance, highlighting their capability to accurately identify tumor regions while maintaining computational efficiency. Key findings include the superior performance of specific EfficientNet variants, which achieved high accuracy and robust segmentation results with reduced training time and resource consumption. This balance between accuracy and efficiency underscores the potential of these models for practical applications in clinical settings, where quick and reliable tumor segmentation is critical for early diagnosis and treatment planning.

The study demonstrates the effectiveness of the EfficientNet family (B0–B7) in segmenting kidney tumors in CT images, leveraging high-performance metrics such as Intersection over Union (IoU) and Dice similarity coefficients. Below is an in-depth analysis of the results and their implications Table 6.1.

Table 6.1: Performance Metrics and Trends.

Metric	Range Observed	Best Model
Mean IoU	0.974 (B1) – 0.980 (B4, B7)	EfficientNet-B4
Mean Dice	0.987 (B1) – 0.990 (B4)	EfficientNet-B4
Background IoU	Consistently 0.999	All Models
Kidney IoU	0.963 (B1) – 0.977 (B7)	EfficientNet-B7
Tumor IoU	0.960 (B0, B3) – 0.968 (B4)	EfficientNet-B4
Background Dice	Consistently 0.999	All Models
Kidney Dice	0.981 (B1) – 0.988 (B7)	EfficientNet-B7
Tumor Dice	0.980 (B0, B3) – 0.984 (B4)	EfficientNet-B4

6.2.1.1 Key Observations

The EfficientNet models demonstrated exceptional segmentation capabilities for kidney tumors in CT images due to their ability to effectively extract global features. Unlike patch-wise classification methods, these models leveraged full-image classification to capture intricate structures such as tissue organization and nuclei patterns, thereby enhancing accuracy in complex scenarios. Background segmentation consistently achieved near-perfect scores, with an IoU and Dice similarity coefficient of 0.999 across all models, showcasing their robust ability to separate irrelevant regions.

Performance improved as the architecture scaled up from EfficientNet-B0 to B4, with B4 achieving the best results in overall segmentation tasks. EfficientNet-B4 excelled in tumor segmentation, attaining an IoU of 0.968, a Dice score of 0.984, and a Mean Dice of 0.990, making it ideal for applications where tumor detection accuracy is critical. EfficientNet-B7, on the other hand, delivered superior results in kidney segmentation, with an IoU of 0.977 and a Dice score of 0.988, demonstrating its strength in precise kidney identification. Lower-complexity models, such as EfficientNet-B0 to B3, maintained competitive performance while requiring fewer computational resources, making them suitable for resource-constrained environments. For example, EfficientNet-B0 achieved a Mean IoU of 0.976 and a Dice score of 0.988, underscoring its efficiency.

However, the study also highlighted certain limitations and trade-offs. Although higher versions like B4 and B7 delivered enhanced accuracy, their increased computational demands could hinder real-time deployment in low-resource settings.

The table 6.2 summarizes the performance of the best-performing EfficientNet models, highlighting their strengths across different segmentation tasks in kidney tumor CT imaging. EfficientNet-B4 emerged as the top model for tumor segmentation, achieving the highest Tumor Dice and overall Mean Dice scores, signifying its exceptional accuracy

in delineating tumor regions and overall segmentation tasks. Meanwhile, EfficientNet-B7 demonstrated superior performance in kidney segmentation, achieving the highest scores for Kidney IoU and Dice metrics. These results emphasize the specialized capabilities of each model, tailored to specific clinical objectives.

Table 6.2: Summary Table of Best Models.

Best Model	Task	Metric	Score
EfficientNet-B4	Tumor Segmentation	Tumor Dice	0.984
EfficientNet-B4	Overall Performance	Mean Dice	0.990
EfficientNet-B7	Kidney Segmentation	Kidney IoU	0.977
EfficientNet-B7	Kidney Segmentation	Kidney Dice	0.988

In summary, EfficientNet architectures exhibit robust performance in segmenting kidney tumours from CT images, with EfficientNet-B4 offering the best trade-off between accuracy and computational efficiency. These results underline the potential of scalable deep learning models in advancing kidney cancer diagnostics.

6.2.2 Comprehensive Analysis of the Results in the Paper: FPN-SE-ResNet Model for Accurate Diagnosis of Kidney Tumors Using CT Images

Chapter four evaluates the efficacy of SE-ResNet models integrated with Feature Pyramid Network (FPN) architecture for segmenting and classifying kidney and tumor regions in CT images of renal cancer. The evaluation focuses on key performance indicators such as Intersection over Union (IoU), accuracy, F1-score, and model loss. The findings reveal that global feature extraction using SE-ResNet outperformed traditional patch-wise approaches by effectively capturing tissue organization. These results underscore the robustness and accuracy of the proposed SE-ResNet architecture when compared to other methods.

Table 6.3: Performance Evaluation of SE-ResNet Models Using IoU Metrics.

Model	Background IoU	Kidney IoU	Tumor IoU	Mean IoU
Seresnet18	0.999	0.972	0.971	0.981
Seresnet34	0.999	0.972	0.971	0.981
Seresnet50	0.999	0.980	0.984	0.988
Seresnet101	0.999	0.976	0.982	0.986
Seresnet152	0.999	0.979	0.984	0.987

Performance evaluation using IoU scores highlights the overlap between predicted and ground-truth segmentation regions, where higher IoU scores reflect better segmentation performance. Table 6.3 presents the IoU scores achieved by various SE-ResNet models. Among these, SE-ResNet50 emerged as the best-performing model, achieving a mean IoU score of 0.988. Across all models, background segmentation maintained consistently high IoU scores of 0.999, while both kidney and tumor IoU scores exceeded 0.97, demonstrating the effectiveness of the SE-ResNet models in segmenting these critical regions.

6.2.2.1 Key Observations

The evaluation of SE-ResNet models integrated with Feature Pyramid Network (FPN) architecture reveals their efficacy in segmenting CT images of kidney and tumor regions. Among these, SE-ResNet50 emerged as the most effective, achieving the highest Intersection over Union (IoU) score of 0.987, outperforming other variants like SE-ResNet18 and SE-ResNet34, both scoring 0.979. Background segmentation performance was uniformly exceptional across all models, with an IoU of 0.999. Tumor IoU scores reached a peak of 0.984 for SE-ResNet50 and SE-ResNet152, underscoring their robust segmentation capabilities.

The Dice Similarity Coefficient (DSC) further validated segmentation accuracy, with SE-ResNet50 achieving scores of 0.990 for kidney segmentation and 0.992 for tumor segmentation. Other models, such as SE-ResNet152 and SE-ResNet101, also demonstrated high Dice scores, reflecting consistent and precise overlap between predictions and ground truth. Loss metrics confirmed the reliability of the models, with validation loss values around 0.752 to 0.753, indicating minimal prediction errors. SE-ResNet50, SE-ResNet101, and SE-ResNet152 exhibited the lowest loss values, emphasizing their optimization strengths.

A comparative analysis highlighted SE-ResNet50’s superiority over established architectures like U-Net, EfficientNet models, and nnU-Net. For example, SE-ResNet50 achieved a Dice score of 0.992 for tumor segmentation, surpassing nnU-Net’s 0.919 and U-Net EfficientNet-B0’s 0.980. Similarly, SE-ResNet50’s IoU scores for kidney and tumor regions (0.980 and 0.984, respectively) exceeded those of other methods such as Hybrid V-Net (0.913) and FR2PAttU-Net (0.911). The model also outperformed EfficientNet-B7 in both Dice and IoU metrics, demonstrating the advantages of the SE-ResNet and FPN integration.

Despite requiring slightly higher computational resources—SE-ResNet50’s training time per unit was 120,970, compared to EfficientNet-B4’s 114,386—the model’s efficiency in parameter usage was notable, with 29.46 million parameters versus EfficientNet-B7’s 75.05 million. Visual analysis further confirmed SE-ResNet50’s precision, with predictions closely matching ground-truth annotations.

In summary, SE-ResNet models consistently achieved high F1-scores, ranging from 0.989 for SE-ResNet18 and SE-ResNet34 to 0.993 for SE-ResNet50 and SE-ResNet152. These results highlight the robustness of SE-ResNet50, which strikes a balance between segmentation accuracy and computational efficiency. While its computational demands may challenge deployment in resource-limited settings, the model’s ability to effectively capture global features and prevent overfitting solidifies its position as the leading choice for kidney and tumor segmentation tasks. Future research should focus on optimizing computational efficiency and exploring its applicability in practical clinical environments.

6.2.3 Comprehensive Analysis of the Results in the Paper: Accurate Kidney Tumor Semantic Segmentation Using Deep Learning Techniques on CT Images

Chapter five explores the efficacy of the DenseNet backbone, integrated with the LinkNet architecture, for extracting global features and segmenting renal cancer CT images. Performance was evaluated using Mean Intersection over Union (IoU) and F1-score metrics. Table 6.4 presents the performance metrics of DenseNet variations paired with the LinkNet architecture. DenseNet121 demonstrates the highest accuracy, with kidney IoU and tumor IoU scores of 0.962 and 0.954, respectively. Similarly, the kidney Dice and tumor Dice scores reach 0.981 and 0.976. DenseNet169 and DenseNet201 also achieve high scores, indicating robustness across the DenseNet family. These results confirm DenseNet's potential in achieving accurate segmentation of both kidney and tumor regions.

Table 6.4: Performance of DenseNet with LinkNet Using Mean IoU and F1-score.

Model	Kidney IoU Score	Tumor IoU Score	Kidney Dice	Tumor Dice
DenseNet121	0.962	0.954	0.981	0.976
DenseNet169	0.956	0.951	0.978	0.975
DenseNet201	0.951	0.953	0.975	0.976

The findings emphasize the superiority of DenseNet121 in renal cancer segmentation, as evidenced by its consistently higher IoU and Dice scores across various categories. This model stands out for its precision and effectiveness in accurately delineating kidney and tumor regions.

An analysis of model accuracy and loss metrics further reinforces the reliability of DenseNet121. Accuracy reflects the model's ability to make correct predictions, while loss indicates the degree of error between predicted and actual outcomes.

DenseNet models generally show a consistent reduction in loss and steady improvements in accuracy throughout training. Among these, DenseNet121 achieves the highest IoU score of 0.969, making it the optimal choice for segmentation and object detection tasks. Its ability to balance high precision with computational efficiency solidifies its role as the preferred model for these applications.

Table 6.5 presents a comparative analysis of the performance of various backbone architectures integrated with different segmentation frameworks, focusing on their ability to segment kidney and tumor regions. Metrics such as IoU (Intersection over Union) and Dice scores for both kidney and tumor segmentation are used to evaluate the models' effectiveness.

The comparison includes established architectures, such as FPN with SE-ResNet50, and proposed configurations combining U-Net, LinkNet, and FPN with backbones like EfficientNetB4 and DenseNet121. The results highlight variations in segmentation performance across different combinations, providing insights into the strengths and

limitations of each approach. This comprehensive evaluation demonstrates the impact of backbone selection and architectural design on segmentation accuracy and precision.

Table 6.5: Performance Comparison of Backbones with Different Architectures.

Reference	Architecture	Backbone	K-IoU-score	T-IoU-score	K-Dice	T- Dice
[29]	FPN	Seresnet50	0.980	0.984	0.990	0.992
paper 4	U-Net	Seresnet50	0.977	0.971	0.988	0.985
paper 4	LinkNet	Seresnet50	0.968	0.971	0.984	0.985
[30]	U-Net	Efficientnetb4	0.974	0.968	0.987	0.984
paper 4	FPN	Efficientnetb4	0.980	0.974	0.990	0.987
paper 4	LinkNet	Efficientnetb4	0.960	0.955	0.980	0.977
paper 4	LinkNet	Densenet121	0.962	0.954	0.981	0.976
paper 4	U-Net	Densenet121	0.937	0.948	0.967	0.973
paper 4	FPN	Densenet121	0.965	0.969	0.982	0.984

The SeresNet50 backbone with FPN achieves the highest overall scores, but the DenseNet121- and EfficientNet-based models exhibit competitive performance. DenseNet121 combined with FPN outperforms DenseNet121-LinkNet, underscoring the impact of architectural choices on segmentation outcomes.

In summary, This comprehensive analysis demonstrates the superior segmentation capabilities of DenseNet-based LinkNet architectures in renal cancer CT imaging. DenseNet121 emerges as the most effective model, combining high segmentation accuracy with computational efficiency. The comparative insights provided in this study will serve as a foundation for future research in medical image analysis, paving the way for improved diagnostic accuracy and treatment strategies.

This chapter compares the performance of EfficientNet, SE-ResNet, and DenseNet-based architectures for segmenting kidney tumors in CT images, emphasizing their ability to balance accuracy and computational efficiency. EfficientNet models, particularly B4 and B7, excel in overall segmentation tasks, achieving high IoU and Dice similarity coefficients for both kidney and tumor regions. SE-ResNet50, integrated with a Feature Pyramid Network (FPN), demonstrates superior segmentation precision, surpassing alternative architectures in IoU and Dice scores while maintaining efficient parameter usage. DenseNet121, combined with LinkNet, achieves competitive segmentation accuracy with stable training performance, highlighting its robustness.

The findings underline the scalability and versatility of these architectures. While EfficientNet offers resource-constrained solutions, SE-ResNet50 achieves the highest segmentation precision, making it ideal for clinical applications. DenseNet121 provides a balance between computational efficiency and performance, suitable for various diagnostic scenarios. Overall, the chapter establishes benchmarks for renal tumor segmentation and proposes directions for optimizing these models for real-world deployments.

6.3 Conclusion

Chapter three presents a rigorous evaluation of the EfficientNet family (B0–B7) for kidney and tumor segmentation in CT imaging. These models effectively address the computational challenges and accuracy demands in automated kidney cancer diagnosis. EfficientNet-B4 achieved the highest tumor segmentation accuracy, with a Tumor Dice score of 0.984 and Mean Dice score of 0.990, while EfficientNet-B7 excelled in kidney segmentation, achieving a Kidney IoU of 0.977 and Kidney Dice of 0.988. These results emphasize the tailored strengths of individual models, with EfficientNet-B4 specializing in tumor detection and EfficientNet-B7 excelling in kidney delineation.

The study highlights EfficientNet’s ability to balance computational efficiency and segmentation precision, with lower-complexity models (e.g., EfficientNet-B0 to B3) offering viable solutions for resource-limited scenarios without significant sacrifices in performance. Key metrics such as consistent background segmentation accuracy (IoU: 0.999) and high overall mean IoU values (0.974–0.980) underscore the reliability of EfficientNet-based architectures for clinical applications.

Despite the promising outcomes, the study also identifies limitations, including the increased computational demands of higher-capacity models like B4 and B7, which could hinder real-time deployment in resource-constrained environments. These insights establish EfficientNet models as scalable and efficient tools for enhancing diagnostic workflows, paving the way for further research to optimize their real-world applicability.

Chapter four investigates the integration of SE-ResNet models with the Feature Pyramid Network (FPN) for kidney tumor segmentation, demonstrating the superiority of this architecture in achieving precise and reliable results. SE-ResNet50 emerged as the best-performing model, achieving a Mean IoU of 0.988 and Dice scores of 0.992 for tumor segmentation and 0.990 for kidney segmentation. The model’s ability to capture global features and enhance segmentation precision surpasses other SE-ResNet variants and state-of-the-art models like nnU-Net and EfficientNet.

Background segmentation across all SE-ResNet models was near-perfect, with IoU consistently at 0.999, further validating their robustness. The findings emphasize SE-ResNet50’s exceptional accuracy, computational efficiency, and consistent performance in delineating critical anatomical structures. While higher-capacity models, such as SE-ResNet101 and SE-ResNet152, showed comparable Dice scores, SE-ResNet50 offered the best balance of accuracy and efficiency, making it ideal for clinical settings.

The chapter underscores the SE-ResNet architecture’s strengths in integrating contextual and spatial information to achieve precise segmentation. However, the slightly higher computational demands of SE-ResNet models highlight the need for further optimization to enable broader real-world deployment, particularly in resource-constrained environments.

Chapter five provides a comparative evaluation of backbone architectures, focusing on DenseNet, SE-ResNet, and EfficientNet in various segmentation frameworks.

DenseNet121, integrated with LinkNet, emerged as the most effective model for kidney and tumor segmentation, achieving Kidney IoU and Dice scores of 0.962 and 0.981, respectively, and Tumor IoU and Dice scores of 0.954 and 0.976. The model demonstrated consistent accuracy and efficiency across multiple metrics, highlighting its robustness for renal cancer CT imaging tasks.

The study's comparative analysis revealed that while SE-ResNet50 with FPN achieved the highest overall segmentation precision (Tumor Dice: 0.992, Kidney Dice: 0.990), DenseNet121 stood out for its balance of computational efficiency and stable training performance. EfficientNet-B4 and B7 also achieved competitive results, particularly in resource-constrained applications, with EfficientNet-B4 excelling in tumor segmentation and B7 in kidney segmentation.

These findings underscore the impact of backbone and architectural design on segmentation outcomes, providing valuable benchmarks for future advancements in kidney tumor diagnostics. The chapter highlights the scalability, versatility, and clinical potential of these architectures, recommending future research into hybrid approaches that combine the strengths of DenseNet, SE-ResNet, and EfficientNet for optimized performance.

6.4 Contribution to Knowledge

The study showcases the effectiveness of EfficientNet models in CT-based kidney and tumor segmentation, with EfficientNet-B4 and B7 excelling in these areas. The research also highlights the balance between precision and computational efficiency, especially in resource-limited scenarios. The integration of SE-ResNet models with the Feature Pyramid Network (FPN) achieves state-of-the-art segmentation accuracy, particularly with SE-ResNet50. A comparative analysis of DenseNet, SE-ResNet, and EfficientNet backbones reveals DenseNet121's superior efficiency and accuracy, SE-ResNet50's exceptional segmentation precision, and EfficientNet's resource-conscious adaptability. The study establishes benchmarks for segmentation precision, computational trade-offs, and scalability, paving the way for optimized diagnostic tools for clinical and resource-constrained environments.

6.5 Future work

Future research can explore the following directions to further enhance segmentation models:

1. Optimization for Real-Time Applications: Tailoring architectures to reduce computational overhead and latency for deployment in real-time diagnostic tools.
2. Integration with Multimodal Data: Combining CT imaging with other modalities, such as MRI or histopathological data, improves segmentation accuracy and robustness.

3. Generalization Across Datasets: Validating and fine-tuning models on diverse datasets to improve generalization and minimize biases across patient demographics.

4. Incorporation of Explainability: Developing interpretable models to assist clinicians in understanding and verifying segmentation decisions, thereby enhancing trust and usability.

References

- [1] A. Saini, M. Kumar, S. Bhatt, V. Saini, and A. Malik, “Cancer causes and treatments,” *Int. J. Pharm. Sci. Res.*, vol. 11, no. 7, pp. 3121–3134, 2020.
- [2] L. B. da Cruz, J. D. L. Araújo, J. L. Ferreira, J. O. B. Diniz, A. C. Silva, J. D. S. de Almeida, A. C. de Paiva, and M. Gattass, “Kidney segmentation from computed tomography images using deep neural network,” *Computers in Biology and Medicine*, vol. 123, p. 103906, 2020.
- [3] S. H. Rossi, T. Klatte, J. Usher-Smith, and G. D. Stewart, “Epidemiology and screening for renal cancer,” *World journal of urology*, vol. 36, pp. 1341–1353, 2018.
- [4] T. Leong, D. Willis, D. L. Joon, S. Condrón, A. Hui, and S. Y. Ngan, “3d conformal radiotherapy for gastric cancer—results of a comparative planning study,” *Radiotherapy and oncology*, vol. 74, no. 3, pp. 301–306, 2005.
- [5] S. Mishra, S. K. Sharma, J. Yadav, and B. Kasana, “A review on how exactly diuretic drugs are working in our body,” *Journal of Drug Delivery and Therapeutics*, vol. 3, no. 5, pp. 115–120, 2013.
- [6] F. Gómez-Pinilla, “Brain foods: the effects of nutrients on brain function,” *Nature reviews neuroscience*, vol. 9, no. 7, pp. 568–578, 2008.
- [7] N. Jourde-Chiche, F. Fakhouri, L. Dou, J. Bellien, S. Burtey, M. Frimat, P.-A. Jarrot, G. Kaplanski, M. Le Quintrec, V. Pernin *et al.*, “Endothelium structure and function in kidney health and disease,” *Nature Reviews Nephrology*, vol. 15, no. 2, pp. 87–108, 2019.
- [8] J. M. Hollingsworth, D. C. Miller, S. Daignault, and B. K. Hollenbeck, “Rising incidence of small renal masses: a need to reassess treatment effect,” *Journal of the National Cancer Institute*, vol. 98, no. 18, pp. 1331–1334, 2006.
- [9] T. Geethanjali and M. Dinesh, “Ensemble approach for improving kidney tumors segmentation performance on ct images using deep learning models.”
- [10] Y. Xie, X. Ma, H. Li, Y. Gao, L. Gu, L. Chen, and X. Zhang, “Prognostic value of clinical and pathological features in chinese patients with chromophobe renal cell carcinoma: a 10-year single-center study,” *Journal of Cancer*, vol. 8, no. 17, p. 3474, 2017.
- [11] W. Zhao, D. Jiang, J. P. Queralta, and T. 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, p. 100357, 2020.

- [12] L. Lipworth, R. E. Tarone, L. Lund, and J. K. McLaughlin, "Epidemiologic characteristics and risk factors for renal cell cancer," *Clinical epidemiology*, pp. 33–43, 2009.
- [13] A. Abdelrahman and S. Viriri, "Kidney tumor semantic segmentation using deep learning: A survey of state-of-the-art," *Journal of imaging*, vol. 8, no. 3, p. 55, 2022.
- [14] E. Roussel, U. Capitanio, A. Kutikov, E. Oosterwijk, I. Pedrosa, S. P. Rowe, and M. A. Gorin, "Novel imaging methods for renal mass characterization: a collaborative review," *European urology*, vol. 81, no. 5, pp. 476–488, 2022.
- [15] R. Suarez-Ibarrola, M. Basulto-Martinez, A. Heinze, C. Gratzke, and A. Miernik, "Radiomics applications in renal tumor assessment: a comprehensive review of the literature," *Cancers*, vol. 12, no. 6, p. 1387, 2020.
- [16] H.-P. Chan, R. K. Samala, L. M. Hadjiiski, and C. Zhou, "Deep learning in medical image analysis," *Deep learning in medical image analysis: challenges and applications*, pp. 3–21, 2020.
- [17] J. Yanase and E. Triantaphyllou, "A systematic survey of computer-aided diagnosis in medicine: Past and present developments," *Expert Systems with Applications*, vol. 138, p. 112821, 2019.
- [18] M. L. Giger, "Machine learning in medical imaging," *Journal of the American College of Radiology*, vol. 15, no. 3, pp. 512–520, 2018.
- [19] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [20] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: achievements and challenges," *Journal of digital imaging*, vol. 32, pp. 582–596, 2019.
- [21] H. Liu and B. Lang, "Machine learning and deep learning methods for intrusion detection systems: A survey," *applied sciences*, vol. 9, no. 20, p. 4396, 2019.
- [22] A. Hatamizadeh, H. Hosseini, Z. Liu, S. D. Schwartz, and D. Terzopoulos, "Deep dilated convolutional nets for the automatic segmentation of retinal vessels," *arXiv preprint arXiv:1905.12120*, 2019.
- [23] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*. Ieee, 2016, pp. 565–571.
- [24] A. Myronenko and A. Hatamizadeh, "3d kidneys and kidney tumor semantic segmentation using boundary-aware networks," *arXiv preprint arXiv:1909.06684*, 2019.

- [25] F. Lateef and Y. Ruichek, “Survey on semantic segmentation using deep learning techniques,” *Neurocomputing*, vol. 338, pp. 321–348, 2019.
- [26] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew, “A review of semantic segmentation using deep neural networks,” *International journal of multimedia information retrieval*, vol. 7, pp. 87–93, 2018.
- [27] F. Isensee and K. H. Maier-Hein, “An attempt at beating the 3d u-net,” *arXiv preprint arXiv:1908.02182*, 2019.
- [28] X. Liu, L. Song, S. Liu, and Y. Zhang, “A review of deep-learning-based medical image segmentation methods,” *Sustainability*, vol. 13, no. 3, p. 1224, 2021.
- [29] A. Abdelrahman and S. Viriri, “Fpn-se-resnet model for accurate diagnosis of kidney tumors using ct images,” *Applied Sciences*, vol. 13, no. 17, p. 9802, 2023.
- [30] ———, “Efficientnet family u-net models for deep learning semantic segmentation of kidney tumors on ct images,” *Frontiers in Computer Science*, vol. 5, p. 1235622, 2023.