Article

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

Abubaker Abdelrahman [ID] and Serestina Viriri *[ID]

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

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].
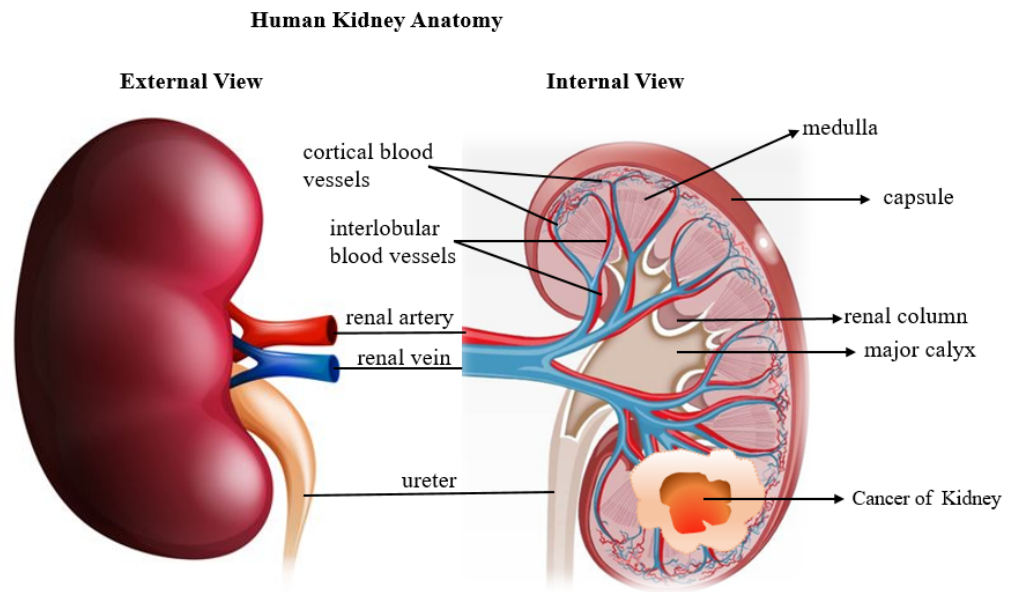
Human Kidney Anatomy

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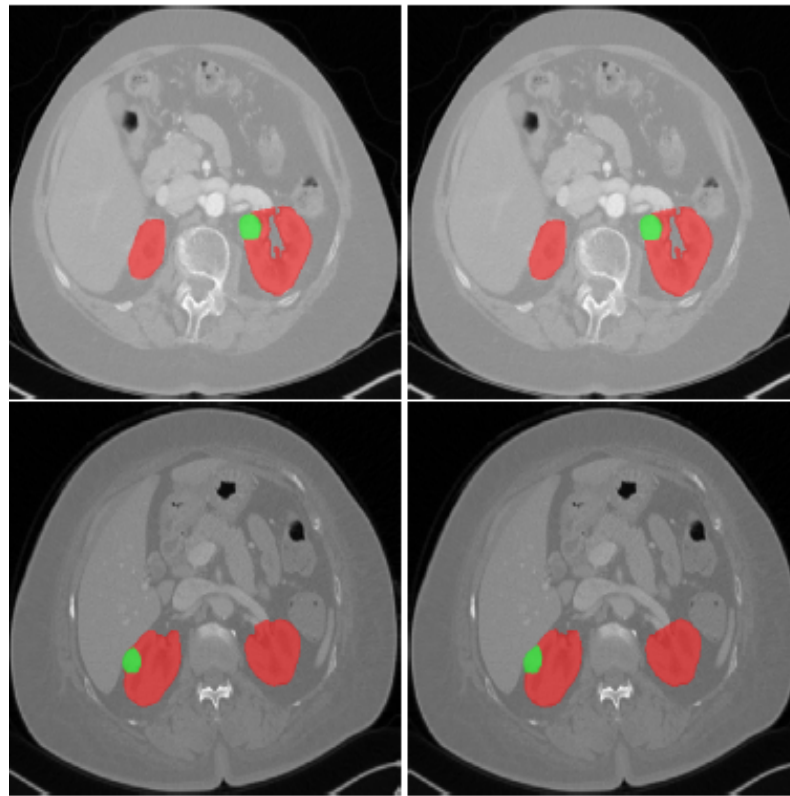
Figure 2. CT scans showing: An axial slice of two patients' 3D CT scans from the KiTS19 dataset; the red tint denotes the kidneys, whereas the green color indicates the tumor site [5].

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, Mark 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 integrationreduces 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 Xxweight1 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 2Xx, and weight2 equals 2 [74].
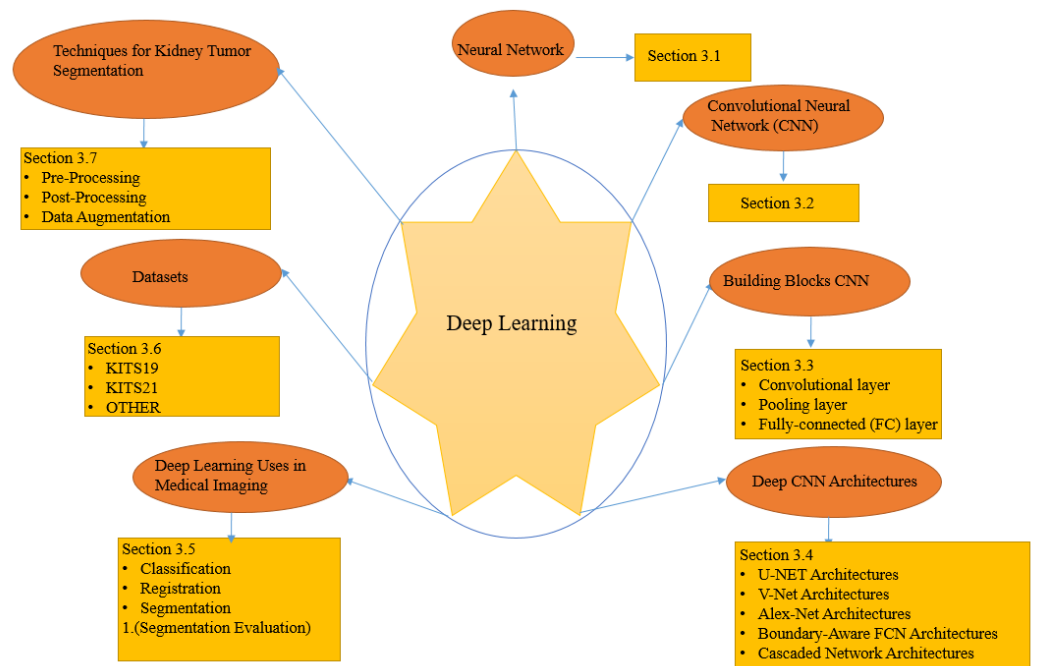
Figure 3. Components architectures, and strategies for deep learning algorithms for segmenting kidney tumors.
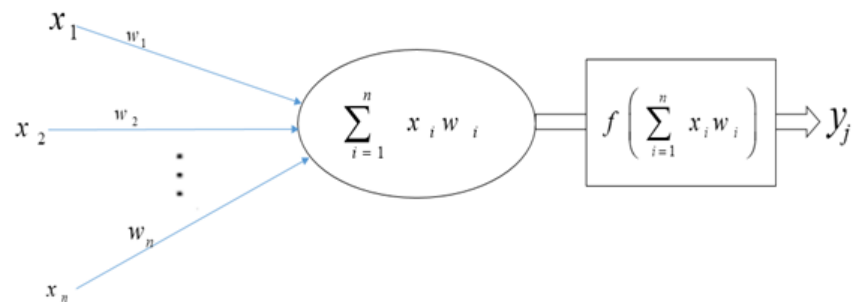

Figure 4. The building block of deep neural networks is an artificial neuron or node.

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.
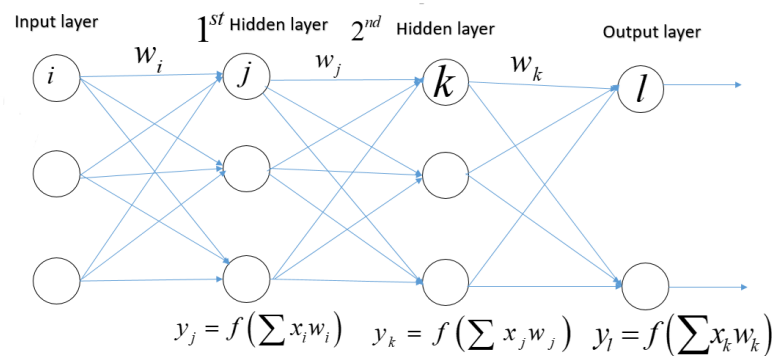
$$y_j = f\left(\sum x_i w_i\right) \quad y_k = f\left(\sum x_j w_j\right) \quad y_l = f\left(\sum x_k w_k\right)$$

Figure 5. Artificial feedforward multilayer neural network.

A feedforward multilayer neural network with two classes (also known as a multilayer perception). 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 HOUP [80], the initial stage, dubbed feature engineering, attempts to extract meaningful data representations from the raw image data. In the second stage, referred known 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.
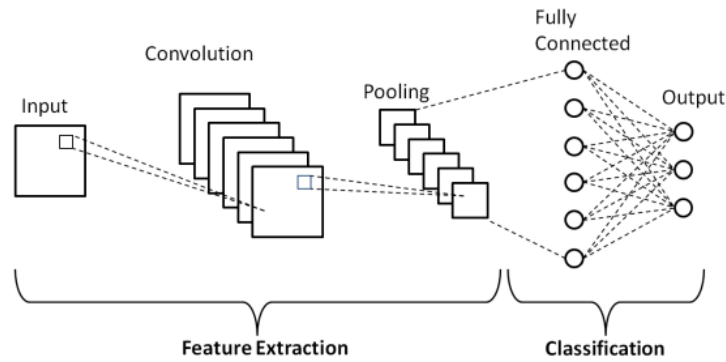

Figure 6. Schematic diagram of a basic convolutional neural network (CNN) architecture [19].

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].

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

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 the 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.
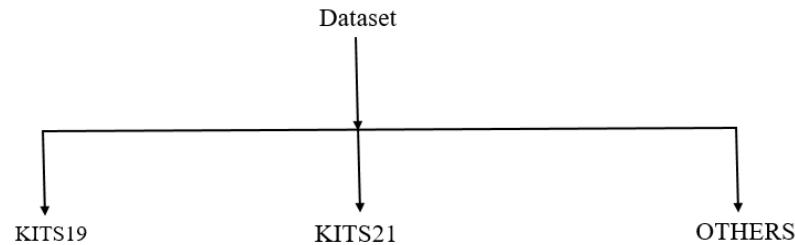

Figure 7. Type of dataset in papers surveyed.

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,

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

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:

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

- 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			Dice	Decathalon
[128]	3D	BN,Depthwise, Weight Pruning	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	ReLUs	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]				Dice	Adam
[19]	3D			Dice	
Ensemble Architecture					
[47]	2D		RELU	Dice	
[64]				Dice	Adam

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 Resent, 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	
[49]	0.975	0.881	0.871
[53]	0.923	0.553	
[65]	0.934	0.643	
[67]	0.96	0.81	
[68]			0.654
[69]	0.916	0.541	
[55]	0.937	0.750	825
[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.

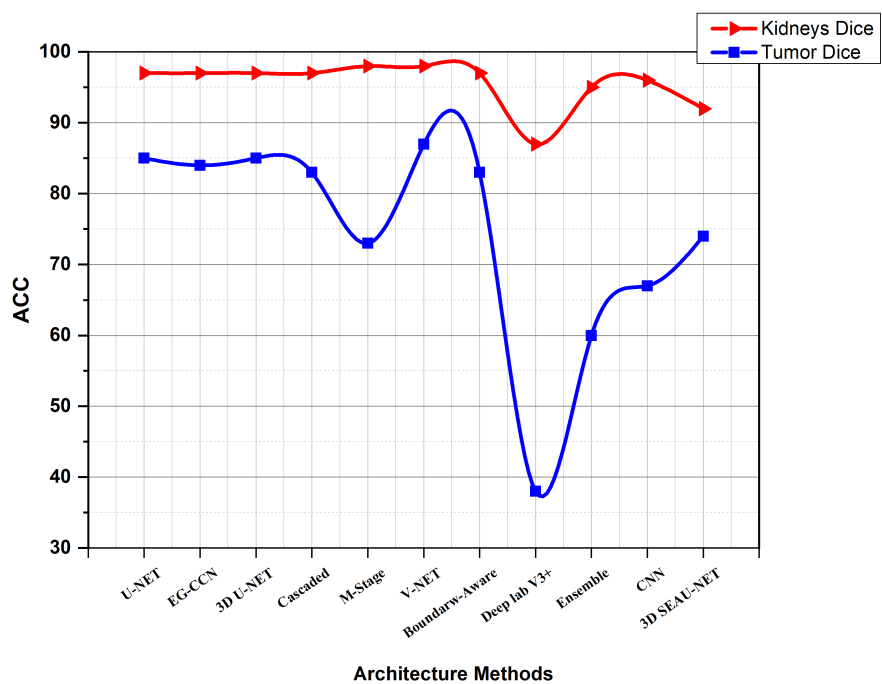| Reference	Sensitivity	Specificity	Jaccard	Accuracy	Hausdorff
[12]			0.716	0.99	33.469
[70]	0.913	0.914			5.10
[50]	0.862	0.894		0.957	
[72]	0.842	0.998	0.756	0.997	18.39

Figure 8. Diagram showing a comparison between different Architecture Methods using KiTs19.
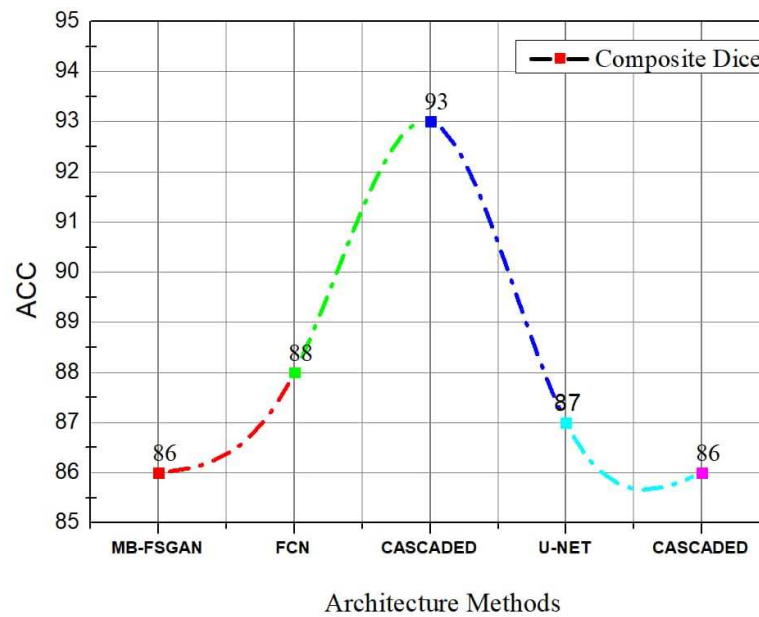
Figure 9. Diagram showing a comparison between different Architecture Methods using KiTs21 and another dataset.

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.

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.; Sathianathen, 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. Greef, 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.; Barışçı, 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=voteINyy36u (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_asJCl (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.; Kwoh, 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-Samplingmethod Based on 3DU-Net Architecture to Reduce the Influence of False Positive and Solve Boundary Blur Problemin 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.; Grimbergen, 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]