FMDB *Transactions on Sustainable Health Science Letters*

Domain-Adaptive Multi-Stage Hybrid Segmentation with Deep Learning and Watershed Techniques for Accurate Medical Image Analysis

Vishal Kumar Kanaujia[1,*], Awadhesh Kumar[2], Satya Prakash Yadav[3]

[1]Department of Computer Science and Engineering-Data Science, ABES Engineering College, Ghaziabad, Uttar Pradesh, India.
[2]Department of Computer Science and Engineering, Kamla Nehru Institute of Technology, Sultanpur, Uttar Pradesh, India.
[3]Department of Computer Science and Engineering, Madan Mohan Malaviya University of Technology, Gorakhpur, Uttar Pradesh, India.
vishalkanaujia.cs@gmail.com[1], awadhesh@knit.ac.in[2], prakashyadav.satya@gmail.com[3]

Abstract: This paper introduces a unique framework for kidney tumour segmentation that leverages deep learning and traditional image processing methods, focusing on the KiTS19 dataset. The proposed approach combines the U-Net architecture with the watershed and random walker algorithms to enhance boundary refinement and segmentation accuracy. To address domain adaptation issues and ensure reliable performance across datasets, a Gradient Reversal Layer (GRL) is also included. The hybrid segmentation model achieves 94.35% accuracy, 93.10% F1-score, and 89.50% Intersection over Union (IoU) on the source domain, demonstrating better performance than conventional deep learning architectures. On unseen target data, the model retains a high accuracy of 91.25%, highlighting the efficacy of domain adaptation strategies. The suggested model has better generalisation and boundary-refinement capabilities than current methods such as U-Net and Fully Convolutional Networks (FCN), making it a good fit for clinical applications that demand accurate segmentation. This multi-stage hybrid technique bridges the gap between deep learning and traditional image processing by improving generalisation performance and ensuring improved border delineation. The findings validate the importance of combining deep learning with domain-adaptation strategies to achieve accurate, generalizable segmentation solutions.

Keywords: Kidney Segmentation; Image Processing; Kidney Tumour Segmentation; Fully Convolutional Networks (FCN); Kidney Cancer; Refinement and Generalisation.

Received on: 02/03/2025, **Revised on:** 05/05/2025, **Accepted on:** 23/07/2025, **Published on:** 08/03/2026

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSHSL

DOI: https://doi.org/10.69888/FTSHSL.2026.000592

Cite as: V. K. Kanaujia, A. Kumar, and S. P. Yadav, "Domain-Adaptive Multi-Stage Hybrid Segmentation with Deep Learning and Watershed Techniques for Accurate Medical Image Analysis," *FMDB Transactions on Sustainable Health Science Letters*, vol. 4, no. 1, pp. 12–30, 2026.

1. Introduction

In medical imaging, precise segmentation of anatomical features and diseased regions, including tumours, is essential for disease monitoring, treatment planning, and diagnosis. However, segmenting kidney tumours poses significant challenges due to their varying sizes, shapes, and heterogeneous appearances. Additionally, the proximity of tumours to complex surrounding

*Corresponding author.

tissues adds further difficulty. This study proposes a multi-stage hybrid segmentation approach that integrates deep learning models with classical segmentation techniques to overcome these challenges and achieve improved boundary refinement and generalisation across datasets. In 2020, more than 430,000 new cases of kidney cancer were reported, making it one of the top 10 most frequent cancers worldwide [1]. Early identification and accurate segmentation of kidney tumours from computed tomography (CT) scans are critical for effective treatment. Although CT imaging provides high-resolution data, manually segmenting tumours is time-consuming and prone to variation across radiologists' skill levels. For clinical applications, automated segmentation techniques are therefore crucial to ensuring precise, reliable, and repeatable findings [13].

1.1. Hybrid Techniques in Medical Image Segmentation

Medical image processing has been greatly enhanced by deep learning, and a key component of many tasks is the convolutional neural network (CNN): classification, detection, and segmentation [2]. However, deep learning models themselves may be insufficient to capture fine tumour boundaries, especially in small and irregular structures. This limitation has motivated the development of hybrid approaches that combine deep learning-based segmentation with classical image processing methods to achieve enhanced performance. Classical algorithms such as watershed segmentation and random walker algorithms excel in refining segmentation boundaries, particularly in regions with connected or overlapping objects [3]. On the contrary, U-Net-based architectures have been acknowledged for their ability to learn hierarchical features and spatial information through encoder-decoder structures with skip connections [4]. Adding classical methods to deep learning workflows helps address boundary issues that may affect tumour segmentation precision [32].

1.2. Domain Adaptation Challenges and Solutions

Another critical challenge in medical segmentation is ensuring generalisation across different datasets—a concept known as domain adaptation. When models trained on one dataset (the source domain) are applied to other datasets (target domains) with varying imaging protocols and qualities, performance tends to degrade [34]. Due to such differences in domain-specific characteristics, conventional deep learning models such as U-Net and Fully Convolutional Networks (FCN) occasionally fail to maintain accuracy on out-of-domain data [5]. In addition, the proposed model will incorporate domain adaptation strategies. Among these methods, the Gradient Reversal Layer (GRL) confuses the domain classifier, thereby improving performance in both source and destination domains, helping the feature extractor learn domain-invariant features [6]. This combination of multi-stage hybrid segmentation and domain adaptation techniques ensures that the model achieves reliable segmentation results, even on datasets with diverse imaging characteristics [28].

1.3. Problem Statement and Objectives

The significant variability of kidney tumours makes it difficult for segmentation models to generalise. By directly learning hierarchical features from raw pixel data, existing deep learning models like U-Net and FCN provide reliable solutions [7]; [8]. However, these models are often limited in their boundary refinement and struggle to generalise across domains without additional techniques. Furthermore, manual segmentation remains burdensome, underscoring the need for an automated framework that integrates deep learning with classical methods to improve precision. This study introduces a multi-stage hybrid segmentation approach that leverages:

- A U-Net-inspired architecture for feature extraction and segmentation.
- Watershed segmentation and random walker algorithms for boundary refinement [9].
- Domain adaptation via Gradient Reversal Layer (GRL) to enhance generalisation across datasets [10].

1.4. Contributions of this Study

The primary contributions of this study are:

- Proposing a multi-stage hybrid segmentation approach that combines deep learning with classical segmentation techniques for improved tumour segmentation [14].
- Employing domain adaptation techniques to enhance the models' capacity to generalise across datasets with varying imaging characteristics.
- The proposed model was assessed using metrics including recall, precision, accuracy, intersection over union (IoU), and F1-score. KiTS19 was used as the dataset, and the results outperformed earlier methods [37].

This study presents a novel kidney tumour segmentation framework that combinsdeep learning and conventional methods to guarantee precise boundary detection [30]. Also, the model performs well across a variety of datasets by incorporating domain adaptation, making it well-suited for practical clinical applications [19].

2. Literature Review

Deep Learning (DL) algorithms have recently become increasingly effective tools for medical image segmentation, demonstrating impressive results across a variety of tasks, including anatomical structure delineation, lesion identification, and organ and tumour segmentation. This section reviews previous studies that provide the foundation for understanding the application of U-Net, Fully Convolutional Networks (FCN), and other DL models to the medical image segmentation problem.

2.1. U-Net in Medical Image Segmentation

The U-Net architecture, introduced by Ronneberger et al. [2], which, through its ability to learn hierarchical features while preserving spatial resolution via skip connections, has become one of the most widely used models for segmenting medical images. Since its inception, many studies have surveyed its application across all domains within medicine. For instance, Isensee et al. [4] presented a self-configuring U-Net that achieved cutting-edge results on a variety of scientific datasets, including liver and brain tumour segmentation tasks. Multi-modal medical image segmentation heavily relies on the U-Net architecture [17]. Tie et al. [11] developed a variant of the U-Net for segmenting brain tumours from MRI and CT data, thereby demonstrating its versatility across imaging modalities. Similarly, prostate segmentation from MRI images performs better when the U-Net is extended to incorporate attention mechanisms [12]. The greater flexibility of U-Net was further demonstrated by Zhou et al. [5] with the proposal of the UNet++ architecture, a nested U-Net specifically designed to improve segmentation accuracy by refining the segmentation process across multiple stages. This architecture has been widely applied to tasks such as liver, cardiac, and kidney tumour segmentation, where the accurate delineation of tumours is critical [18] (Table 1).

Table 1: U-Net in medical image segmentation

Study	Focus Area	Contribution	Application
Ronneberger et al. [2]	U-Net Architecture	Introduced the U-Net for medical image segmentation	General medical image segmentation
Isensee et al. [4]	nnU-Net	Proposed a self-configuring version of U-Net with state-of-the-art performance	Liver and brain tumor segmentation
Tie et al. [11]	Multi-modal Segmentation	Used U-Net to segment brain tumours from MRI and CT scans	Brain tumor segmentation
Hong et al. [12]	Attention Mechanisms	Integrated attention mechanisms into the U-Net to improve segmentation accuracy	Prostate segmentation
Zhou et al. [5]	U-Net++	Introduced a nested U-Net architecture to enhance segmentation accuracy	Liver, cardiac, and kidney tumour segmentation

2.2. Fully Convolutional Networks (FCN) and their Variants

Long et al. [3] introduced Fully Convolutional Networks (FCN), one of the earliest deep learning models for image segmentation, as the first CNN design to replace convolutional layers with fully connected layers for dense, pixel-wise prediction. Since then, FCN has been applied to several medical imaging tasks, including lung nodule recognition and retinal image segmentation [21]. However, the model could not capture many fine details; thus, improvements and alternatives were explored. Recent studies have mainly focused on improving FCN's performance through architectural changes [25]. Using architectures specifically designed for 3D medical image segmentation applications, such as prostate cancer segmentation, Milletari et al. [15] developed V-Net, for instance, as a volumetric extension of FCN. Similarly, Dou et al. [16] demonstrated 3D FCN designs and achieved better results in volumetric medical data segmentation, including liver and pancreatic segmentation. Numerous studies look at attention-based FCN models. To improve segmentation performance by focusing on the most important factors, Oktay et al. [7] introduced the Attention U-Net, which integrates an FCN with attention mechanisms. Tasks such as liver tumour segmentation and cardiac image segmentation have used this architecture (Table 2).

Table 2: Fully convolutional networks (FCN) and their variants

Study	Focus Area	Contribution	Application
Long et al. [3]	FCN Architecture	Replaced fully connected layers with convolutional layers	Early image segmentation tasks
Milletari et al. [15]	V-Net	Introduced a 3D extension of FCN for volumetric data	Prostate cancer segmentation

Dou et al. [16]	3D FCN	Utilised dense blocks for improved volumetric segmentation	Pancreas and liver segmentation
Oktay et al. [7]	Attention U-Net	Combined FCN with attention mechanisms for enhanced segmentation	Cardiac and liver tumor segmentation

2.3. Comparative Studies of U-Net and FCN

Several comparative studies have assessed the relative advantages and disadvantages of U-Net and FCN in medical picture segmentation [26]. For instance, compared with Christ et al. [20], who analysed the two models for liver and tumour segmentation, the authors found that FCN offered better computational efficiency, whereas U-Net performed well in boundary correctness. Similarly, researchers compared U-Net and FCN for brain tumour segmentation and observed that U-Net achieved higher Dice scores, but FCN demonstrated better generalisation on unseen test cases. Recent work by Mak [21] examined how well U-Net and FCN performed in kidney tumour segmentation using the KITS19 dataset, the same dataset used in this study. Their results corroborate previous research, showing that U-Net's skip connections help retain spatial information, leading to more accurate tumour boundary delineation. At the same time, FCN is more computationally efficient and thus faster to train (Table 3).

Table 3: Comparative studies of U-Net and FCN

Study	Comparison Focus	Key Findings	Application
Christ et al. [20]	U-Net vs. FCN	U-Net showed better boundary accuracy, while FCN was more computationally efficient.	Liver and tumor segmentation
Chen et al. [29]	U-Net vs. FCN	U-Net achieved higher Dice scores, whereas FCN demonstrated better generalisation.	Brain tumor segmentation
Mak [21]	U-Net vs. FCN	U-Net excelled in tumour boundary delineation, while FCN trained faster	Kidney tumor segmentation

2.4. Attention Mechanisms and Hybrid Architectures

To further enhance segmentation performance, attention mechanisms have been increasingly integrated into U-Net and FCN architectures [31]. For example, Schlemper et al. [23] introduced Attention-Gated Networks (AG-Net), which adaptively highlight important regions in medical images during segmentation, improving the network's focus on tumours or lesions. AG-Net has been successfully applied to retinal, cardiac, and lung nodule segmentation [37]. In several other architectures, hybrid designs combine the benefits of multiple models [36]. For example, Du et al. [27] proposed a hybrid model combining FCN and U-Net to leverage the benefits of both architectures. In tasks such as liver segmentation, where it is essential to capture both fine-grained details and large-scale context, this hybrid approach has been shown to improve segmentation performance (Table 4).

Table 4: Attention mechanisms and hybrid architectures

Study	Focus Area	Contribution	Application
Schlemper et al. [23]	AG-Net	Introduced attention-gated networks to highlight important regions	Retinal, cardiac, and lung nodule segmentation
Du et al. [27]	Hybrid U-Net + FCN	Combined U-Net and FCN to capture both global context and fine details	Liver segmentation

2.5. Other Advances in Medical Image Segmentation

In addition to U-Net and FCN, other deep learning models have also shown impressive performance on the segmentation task for medical images [24]. For example, DeepLab increases the receptive field by doubling the number of atrous convolutions to prevent the number of parameters from rising [33]. This approach has been effectively used for tasks such as lung cancer detection and brain tumour segmentation. Zhao et al. [35]'s pyramid scene parsing network is another remarkable architecture that captures both local and global contextual information. It has been demonstrated that PSPNet outperforms other models on tasks such as lesion identification and organ segmentation [35]. Medical image segmentation trends now are toward integration with 3D architectures and the use of transfer learning. To this end, Li et al. [37] investigated the application of the 3D U-Net for segmenting organs in volumetric data, such as the liver and pancreas, from CT images. Similar gains in efficiency are due to the applicability of transfer learning with pre-trained models and to fine-tuning networks such as U-Net on medical datasets with limited labelled data (Table 5).

Table 5: Other advances in medical image segmentation

Study	Focus Area	Contribution	Application
Chen et al. [29]	DeepLab	Introduced atrous convolutions to enlarge receptive fields	Brain tumour and lung cancer detection
Zhao et al. [35]	PSPNet	Captured both local and global contextual information using pyramid pooling	Organ segmentation and lesion detection

3. Material and Method

3.1. Dataset

For kidney tumour segmentation, the KITS19 dataset comprises annotated pictures. To preprocess the photographs, researchers resized them to 256x256, normalised the pixel values, and added information by flipping and rotating the images.

3.2. Model Architectures

3.2.1. U-Net

FCN was specifically designed for biomedical image segmentation and has shown remarkable effectiveness in applications that require accurate localisation, such as organ and tumour segmentation. An encoder-decoder design serves as its foundation, incorporating symmetric skip connections between matching layers along the encoder and decoder paths. By employing skip connections, the network can recover spatial information that is skipped at each downsampling step, thereby maintaining high-frequency properties essential for detecting sharp borders. Thus, the architecture is separated into two sections:

- **Encoder (Contracting Path):** The encoder enhances feature encoding while simultaneously reducing the spatial resolution of the supplied image by layering several convolutional and pooling layers. After each downsampling step, two 3x3 convolution layers without padding, ReLU activations, and a 2x2 max pooling operation are applied. There are twice as many feature channels in each downsampling step.
- **Decoder (Expanding Path):** The decoder iteratively improves spatial resolution through a series of convolutional and upsampling operations. Each feature map is scaled up in the decoder using a 2x2 deconvolution (also known as "up-convolution"), and then inserted using skip connections from the encoder to the corresponding feature map. This enables the model to leverage both high-level and low-level information, thereby improving its localisation performance in segmentation tasks.
- **Mathematical Model:** Let the input image be denoted by $X \in R^{H\times W\times C}$, where H = height, W =width, and C = number of channels (e.g., 1 for grayscale, 3 for RGB). The segmentation mask is denoted $Y \in \{0,1\}$ ^H×W×K, where K denotes the number of output classes (typically K = 2 for binary segmentation). For each layer l in the encoder, the output feature maps f_l are computed using convolutional filters $W_l \in R^{kxkxFl}$ (where k is the kernel size and F_l defines the number of feature maps at l layer l), biases b_l, and the ReLU activation function $\sigma(\cdot)$. This operation can be described as:

$$f_{l+1} = \sigma(W_l * f_l + b_l)$$

Where the convolution process is indicated by $*$ and f_l is the output feature map of layer l. For the pooling operation, a max-pooling function reduces the spatial resolution by a factor of 2. In the decoder path, the transposed convolution (or up-convolution) W_{up} is applied to upsample the feature map f_l. Mathematically, this can be written as:

$$f_{up} = W_{up} *^T f_l$$

Where $*^T$ represents the transposed convolution operation, this adds a factor of two to the feature map's spatial dimensions. The skip connections are formulated as concatenation operations, where the output of the encoder layer f_l^{enc} is concatenated with the output of the corresponding decoder layer f_l^{dec}:

$$f_l^{skip} = [f_l^{enc}, f_l^{dec}]$$

Additional convolutional layers are subsequently applied to the concatenated feature map to improve the segmentation result:

- **Final Segmentation:** After passing through the encoder-decoder network, the final segmentation map $\hat{Y} \in R^{H\times W\times K}$ is generated using a 1x1 convolution applied to the final feature map:

$\hat{Y} = \sigma(W_{final} * f_{final} + b_{final})$

Where $W_{final} \in R^{1\times 1\times F_{final}\times K}$ is the 1x1 convolutional filter that filters the feature map channels down to K classes, and for binary segmentation, $\sigma(\cdot)$ is often a sigmoid function; for multiclass segmentation, it is a softmax function:

- **Loss Function:** For binary segmentation jobs, the binary cross-entropy loss is frequently used and is defined as follows:

$$\mathcal{L}_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\left[Y_i \log(\hat{Y}_i) + (1 - Y_i)\log(1 - \hat{Y}_i)\right]$$

Where N = total number of pixels, Y_i = true label for pixel i, and $\hat{Y}_i$ is the predicted probability for pixel i. It is possible to use a categorical cross-entropy loss for multi-class segmentation tasks:

$$\mathcal{L}_{CCE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} Y_{ik}\log(\hat{Y}_{ik})$$

Where K denotes the number of classes, and Y_{ik} denotes the true label for class k at pixel i. This architecture's ability to combine coarse and fine features via skip connections makes the U-Net particularly well-suited for biomedical segmentation tasks, where detailed boundaries are critical (Table 6).

Table 6: U-Net configuration

Layer Type	Input Size	Filter Size	Output Channels	Stride	Activation
Input Layer	256x256x1	-	1	-	-
Conv2D + ReLU	256x256x1	3x3	64	1	ReLU
Conv2D + ReLU	256x256x64	3x3	64	1	ReLU
MaxPooling2D	256x256x64	2x2	64	2	-
Conv2D + ReLU	128x128x64	3x3	128	1	ReLU
Conv2D + ReLU	128x128x128	3x3	128	1	ReLU
MaxPooling2D	128x128x128	2x2	128	2	-
(Encoder continues...)	...	...	...	...	...
Transposed Conv2D	16x16x1024	2x2	512	2	ReLU
Conv2D + ReLU	32x32x1024	3x3	512	1	ReLU
Conv2D + ReLU	32x32x512	3x3	512	1	ReLU
(Decoder continues...)	...	...	...	...	...
Output Layer	256x256x64	1x1	2	-	Softmax

3.2.2. Fully Convolutional Network (FCN)

Another architecture used for segmentation is the Fully Convolutional Network (FCN); however, it differs from U-Net in its structure, lacking symmetric skip connections. Convolutional layers, commonly used in classification networks, replace fully connected layers in FCNs, enabling the network to produce dense pixel-level predictions. The model involves an encoder that extracts hierarchical features from the input image, followed by subsequent upsampling (transposed convolution) layers for restoring resolution:

- **Mathematical Model:** Let $X \in R^{H\times W\times C}$ be the input image. The output feature maps at layer l are computed as:

$f_{l+1} = \sigma(W_l * f_l + b_l)$

Similar to U-Net, W_l represents the convolutional filters, b_l represents the bias, and $*$ denotes the convolution operation. In the decoder, the upsampling is performed using transposed convolutions, similar to U-Net:

$f_{up} = W_{up} *T\ fl$

Where $*^T$ represents the transposed convolution operation to increase the spatial resolution (Table 7).

Table 7: FCN configuration

Layer Type	Input Size	Filter Size	Output Channels	Stride	Activation
Input Layer	256x256x3	-	3	-	-
Conv2D + ReLU	256x256x3	3x3	64	1	ReLU
Conv2D + ReLU	256x256x64	3x3	64	1	ReLU
MaxPooling2D	256x256x64	2x2	64	2	-
Conv2D + ReLU	128x128x64	3x3	128	1	ReLU
Conv2D + ReLU	128x128x128	3x3	128	1	ReLU
MaxPooling2D	128x128x128	2x2	128	2	-
(Encoder continues...)	...	...	...	...	...
Conv2D + ReLU	32x32x512	1x1	512	1	ReLU
Transposed Conv2D	32x32x512	4x4	256	2	ReLU
Skip Connections	64x64x256	-	-	-	-
(Decoder continues...)	...	...	...	...	...
Output Layer	256x256x512	1x1	21	-	Softmax

Then, a 1x1 convolution maps the last feature maps to the number of classes, and the output passes through the softmax or sigmoid activation function, depending on whether the task is binary or multi-class segmentation (Table 8).

Table 8: Layer-wise breakdown of the proposed model architecture

Layer (type)	Output Shape	Param #	Connected to
input layer (InputLayer)	(None, 64, 64, 3)	0	-
conv2d (Conv2D)	(None, 64, 64, 32)	896	input layer[0][0]
max pooling2d (MaxPooling2D)	(None, 32, 32, 32)	0	conv2d[0][0]
conv2d 1 (Conv2D)	(None, 32, 32, 64)	18,496	max pooling2d[0][0]
max pooling2d 1 (MaxPooling2D)	(None, 16, 16, 64)	0	conv2d 1[0][0]
conv2d 2 (Conv2D)	(None, 16, 16, 128)	73,856	max pooling2d 1[0][0]
max pooling2d 2 (MaxPooling2D)	(None, 8, 8, 128)	0	conv2d 2[0][0]
up sampling2d (UpSampling2D)	(None, 16, 16, 128)	0	max pooling2d _2[0][0]
conv2d 3 (Conv2D)	(None, 16, 16, 64)	73,792	up sampling2d[0][0]
up sampling2d _1 (UpSampling2D)	(None, 32, 32, 64)	0	conv2d 3[0][0]
flatten (Flatten)	(None, 8192)	0	max pooling2d 2[0][0]
conv2d 4 (Conv2D)	(None, 32, 32, 32)	18,464	up sampling2d 1[0][0]
gradient reversal _layer (Gradient Reversal Layer)	(None, 8192)	0	flatten[0][0]
up sampling2d _2 (UpSampling2D)	(None, 64, 64, 32)	0	conv2d 4[0][0]
dense (Dense)	(None, 128)	1,048,704	gradient reversal layer[0][0]
conv2d 5 (Conv2D)	(None, 64, 64, 16)	4,624	up sampling2d 2[0][0]
dropout (Dropout)	(None, 128)	0	dense[0][0]
conv2d 6 (Conv2D)	(None, 64, 64, 1)	17	conv2d 5[0][0]
dense 1 (Dense)	(None, 1)	129	dropout[0][0]

Evaluating the U-Net and FCN using critical metrics such as Dice Similarity Coefficient (DSC), Specificity, and Sensitivity, the two models' capabilities to accurately identify or segment the target structures are compared (Figure 1).

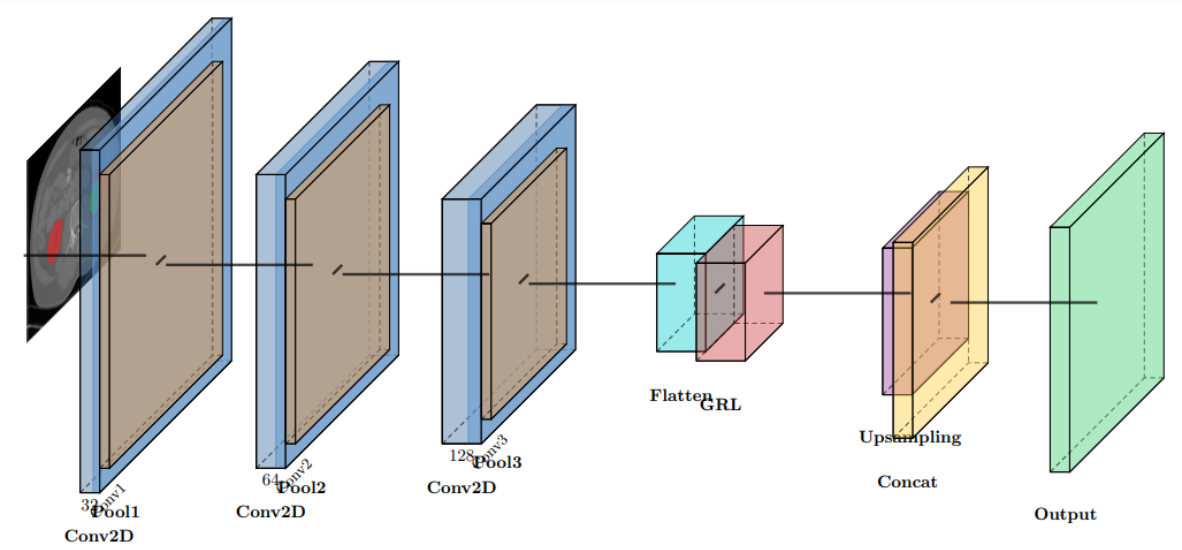

Figure 1: Layered architecture of the proposed model

3.3. Training Setup

The training of both U-Net and FCN models was carried out using a carefully designed setup to ensure optimal learning and convergence. Below are the critical components of the training pipeline:

- **Optimiser:** As it combines the advantages of two well-known optimisers, RMSProp and AdaGrad, researchers used the Adam optimiser, which is a popular choice. Adam uses estimates of the gradients' first and second moments to modify the learning rate for each parameter:

$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$

$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$

Where m_t = estimate of the first (mean) and v_t = estimate of the second (uncentered variance) moments of the gradients, respectively, and g_t = gradient at time step t. The parameters β_1 and β_2 are typically set to 0.9 and 0.999, respectively. The update rule for the weights W is then given by:

$$W_{t+1} = W_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}$$

Where η = learning rate, ϵ =small constant to avoid division by zero, and $\hat{}m_t$ and $\hat{}v_t$ are the bias-corrected estimates of m_t and v_t, this combination of adaptive learning rates and momentum leads to more robust convergence, especially in noisy gradient environments:

- **Learning Rate:** For training, a fixed learning rate of 0.001 was used to balance stability and fast convergence. Gradient descent used in optimisation accounts for the size of the update steps. Here, the loss function's global minimum will not be exceeded during the weight update procedure.
- **Batch Size:** A batch size of 16 was selected to maximise memory usage and ensure reliable gradient estimation. The number of training samples processed before the model's parameters are updated depends on the batch size. Smaller batch sizes, such as 16, increase the number of updates, improving training performance and better utilising the GPU's current parallel computing capabilities.
- **Epochs:** Ten epochs were used to train the models. Each epoch represents a complete pass through the training dataset. In fact, this number can be adjusted based on the validation loss and other early-stopping criteria to further refine the models, even though training for a small number of epochs (e.g., 10) helps avoid overfitting.
- **Loss Function:** In binary segmentation tasks, where the objective is to assign each pixel to one of two classes (e.g., tumour or background), the model's performance was evaluated using the BCE loss. The following provides the BCE loss function:

$$L_{BCE}(Y,\hat{Y}) = -\frac{1}{N}\sum_{i=1}^{N}\left(Y_i \log(\hat{Y}_i) + (1-Y_i)\log(1-\hat{Y}_i)\right)$$

where N is the total number of pixels, $Y_i \in \{0,1\}$ is the true label for pixel i, and $\hat{Y}_i \in [0,1]$ is the predicted probability that pixel i belongs to the class of interest (e.g., tumour). Better performance is indicated by lower BCE loss values, which measure the difference between the expected probability and the actual labels. For multi-class segmentation tasks, this can be expanded to the categorical cross-entropy loss:

- **Regularisation:** To prevent overfitting, L_2 regularisation was applied to the model weights. The regularisation term is added to the loss function as follows:

$$L_{total} = L_{BCE} + \lambda \sum_{l} \|W_l\|_2^2$$

Where λ = regularisation coefficient, and W_l = weights of layer l. L2 regularisation penalises large weights, leading to a smoother model that generalises better to unseen data:

- **Data Augmentation:** To make the models more robust against overfitting and improve their resilience, random rotations, flips, zooms, and shifts were applied to the training data. Data augmentation, by making modified copies of the original images, creates an expanded training set that helps the model learn more invariant characteristics.
- **GPU Acceleration:** Both U-Net and FCN were trained using NVIDIA GPUs, leveraging CUDA for efficient computation. By parallelising the numerous matrix operations (such as convolutions and upsampling) used by the models, it drastically reduces training time. GPUs allow for faster training, especially when working with large datasets and deep architectures.

3.4. Training Process

The training procedure for a deep learning model is iterative and involves multiple cycles of epochs. In each epoch, the data set is split into smaller batches. For each batch, the model goes through a sequence of steps. First, the forward pass consists of feeding the batch input data, Xbatch, into the network, producing predictions, $\hat{Y}_{batch}$. After this, the loss is calculated by comparing the model's predictions to the true labels, Y_{batch}. Loss measures how far the predictions are from the actual values. Next, the model has the backward pass. It calculates the gradients of the loss with respect to its parameters. These gradients guide how to adjust the network's parameters to minimise loss. Finally, it actually updates the parameters with the Adam optimiser. Then the cycle repeats: forward pass and loss, and then backward pass and parameter update on each batch in the dataset, and so forth. Once all the batches in the dataset are processed, one epoch has passed. The process then repeats, beginning with a different epoch, and continues until the specified number of epochs is completed and the model has learnt sufficiently. Additionally, throughout each training iteration, accuracy and validation loss have been tracked to ensure the learned models generalise effectively to new data. Also, early stopping criteria have been used to avoid overfitting: training was halted if the validation loss did not improve after a predetermined number of epochs.

3.4.1. Evaluation Metrics

To evaluate the models, several metrics were used:

- **Dice Similarity Coefficient (DSC):** This assesses the degree of overlap between the ground truth and the predicted segmentation:

$$DSC = \frac{2 \cdot |Y \cap \hat{Y}|}{|Y| + |\hat{Y}|}$$

Where Y= expected mask and $\hat{Y}$= ground truth mask. Higher DSC values indicate better segmentation performance, with values ranging from 0 to 1:

- **Sensitivity (SEN):** Also known as recall. The percentage of true positives (pixels that are correctly detected) among all actual positives is known as recall:

$$SEN = \frac{TP}{TP + FN}$$

Where,

TP = Number of true positives
FN = Number of false negatives

- **Specificity (SPE):** Out of all genuine negatives, it calculates the percentage of true negatives that are accurately detected background pixels:

$$SPE = \frac{TN}{TN + FP}$$

Where,

TN = True Negatives
FP = False Positives

3.5. Training Procedure

3.5.1. Input Image Acquisition

- The segmentation process begins with input images, typically CT or MRI scans in NIfTI format, that contain anatomical and pathological regions.
- These images are processed using the Nibabel library for loading the NIfTI files. In total, these metrics give an all-around evaluation of whether the models really do a good job in the segmentation of the target structures on medical images.

3.5.2. Classical Image Processing Techniques

- When classical methods are used, thresholding is applied to perform basic binary segmentation, separating potential tumour regions from the background.
- Watershed segmentation refines tumour boundaries, while the Random Walker algorithm further separates connected components and enhances segmentation accuracy.

3.5.3. Deep Learning-Based Segmentation

- If a deep learning model is employed, a U-Net or a customised architecture is used.
- Using CT imaging and its corresponding truth masks, the model can predict tumour regions at the pixel level.
- The encoder-decoder design pattern, for example, with skip connections, can effectively capture low-level details and high-level semantics.

3.5.4. Combining Results from Both Approaches

- The outputs of both classical techniques (e.g., Watershed, Random Walker) and the deep learning model (e.g., U-Net) are combined.
- This hybrid approach ensures accurate segmentation by leveraging the strengths of both methodologies—classical methods for boundary refinement and deep learning for robust feature extraction.

3.5.5. Domain Adaptation for Generalisation

- A Gradient Reversal Layer (GRL) is used to train a domain-invariant feature extractor. This ensures the model performs well across multiple datasets or imaging modalities.
- The GRL forces the network to confuse the domain classifier, making the feature representations domain-agnostic.

3.5.6. Handling Domain Shift

- If a domain shift is detected (e.g., when the model faces new imaging data with different characteristics), model parameters are adjusted to maintain performance.
- If no domain shift is detected, the current model configuration is retained.

3.5.7. Model Training and Optimization

- The provided network is trained using the Adam optimiser in conjunction with the binary cross-entropy loss. Optimisation of the pixel-wise binary classification problem is complete.
- Training is performed over multiple epochs, with both accuracy and IoU (Intersection over Union) used for evaluation.

3.5.8. Output Segmented Image

- The final segmented image is generated, highlighting the tumour region within the input image.
- The result is visualised by overlaying the predicted mask on the original scan to assess segmentation quality.

3.5.9. Evaluation of Tumour Detection Accuracy

The model's segmentation performance is assessed using important metrics such as recall, accuracy, precision, F1-score, and IoU. These metrics indicate the model's general abilities across domains and its accuracy in detecting tumour boundaries (Figure 2).

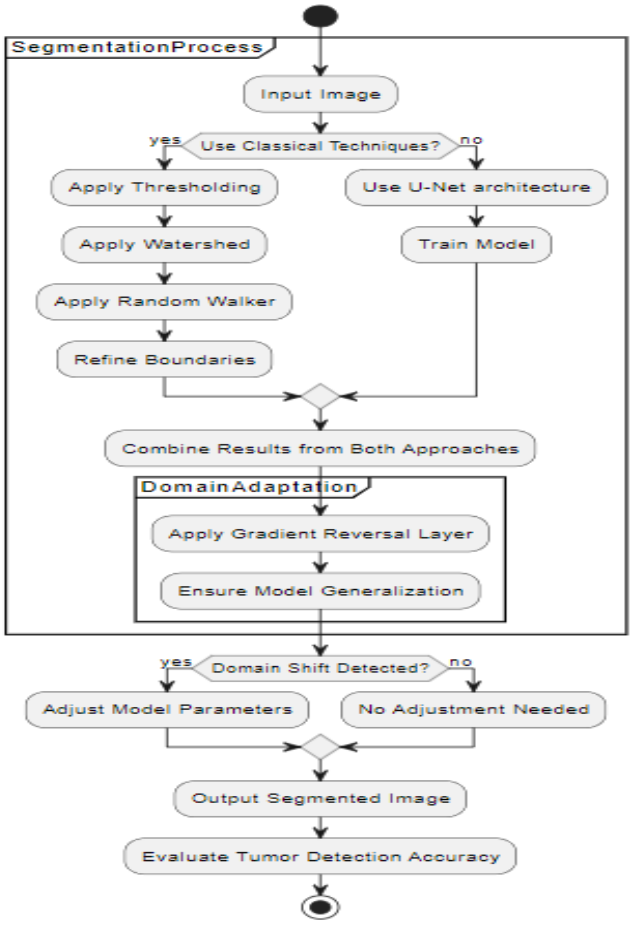

Figure 2: Flowchart of the training procedure

4. Results and Discussions

In addition to achieving a high index of performance, both U-Net and FCN models were evaluated using assessment metrics such as specificity, sensitivity, and the dice similarity coefficient. The outputs of these models are compared and elaborated in the subsequent sections. To understand the results produced by the segmentation models, researchers include images of original medical scans, segmentation ground truths, and tumour mask predictions. These images show, from left to right: Original Imaging, Ground Truth Segmentation of Tumour, and Tumour-Enhanced Imaging. These images also demonstrate the U-Net model's effectiveness at heat-mapping tumour regions using the provided ground truth (Figure 3).

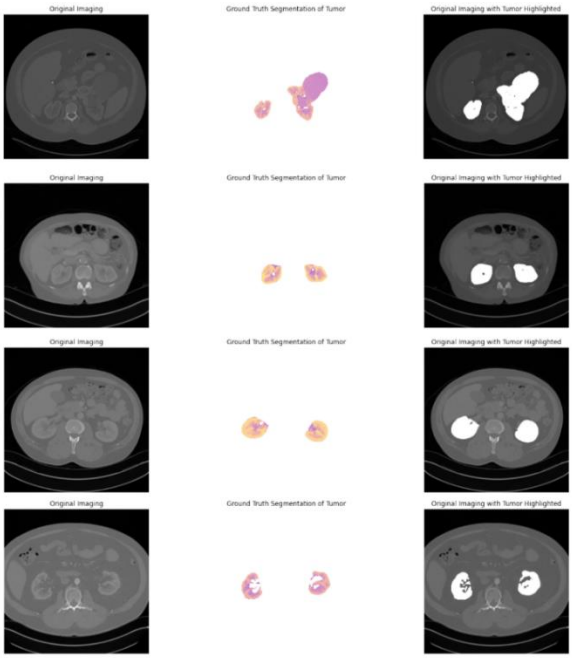

Figure 3: Sample output from the U-Net model

Original Imaging, Binary Segmentation, Ground Truth Segmentation, and Tumour only are listed from left to right. The binary segmentation results highlight the FCN model's effectiveness in delineating tumour regions, though slight discrepancies in boundary accuracy compared with the U-Net are evident (Figure 4).

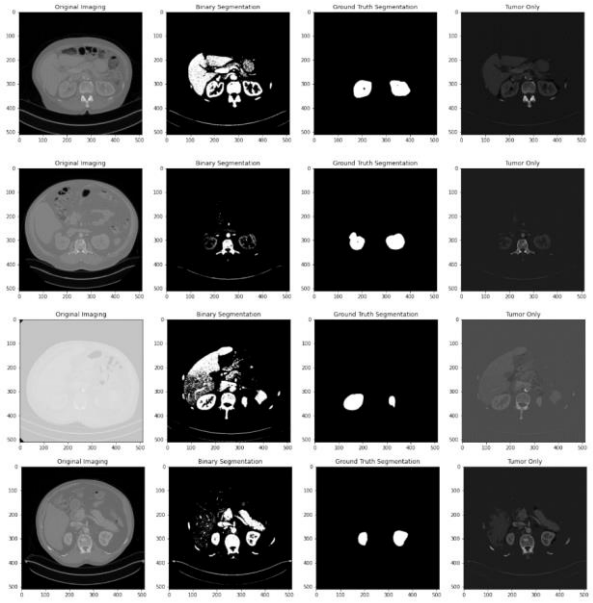

Figure 4: Sample output from the FCN model

The assessment of U-Net and FCN models is based on the evaluation metrics: SPE, SEN, and DSC. These metrics illustrate the performance of each model in segmenting kidney tumours from medical images in the KITS19 dataset. The results presented below provide important insights into the advantages and disadvantages of the models in question.

4.1. Training and Validation Loss

During training, the validation and training losses for the FCN and U-Net models were monitored. As shown in Figure 5 below, the models exhibited different behaviours during optimisation of the loss function.

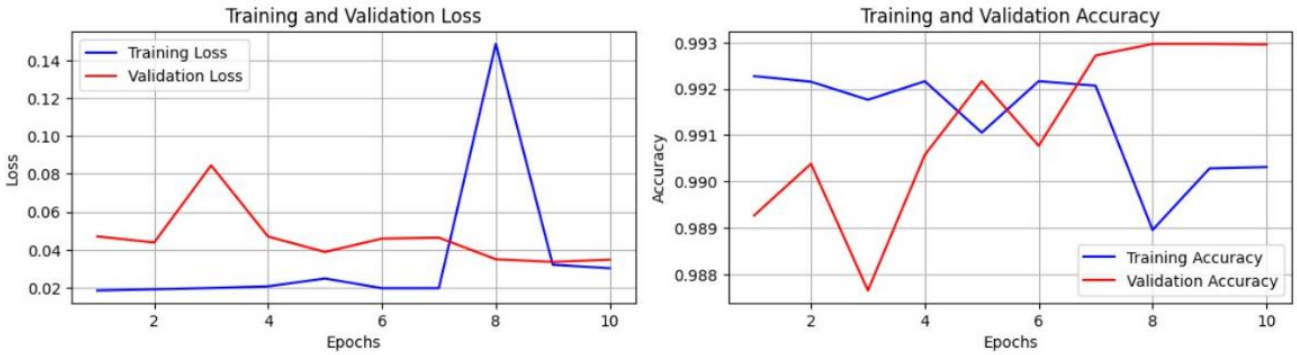

Figure 5: Training and validation loss for U-Net and FCN

- **U-Net:** U-Net demonstrated a relatively smooth learning curve, with stable training loss throughout the epochs. The validation loss, though slightly more erratic at times, also converged to a low value by the 10th epoch. The minor fluctuations in validation loss likely reflect the model's response to the variability in the validation set, but the overall trend indicates good generalisation. This consistency implies that the U-Net is not only learning effectively but also balancing the preservation of accuracy on unseen data with fitting the training data.
- **FCN:** In contrast, the FCN model experienced more variability in the validation loss, particularly during the earlier epochs. This suggests that the FCN model had greater difficulty effectively adjusting its weights early on. However, as training progressed, the FCN model also converged to a minimum loss, indicating eventual stabilisation. The initial fluctuations could be attributed to the model's simpler architecture, which lacks the U-Net's skip connections, making it more sensitive to complex spatial features in the data.

By preserving significant spatial information from the input images, U-Net's architecture — specifically its use of skip connections — offers a more stable learning process, as indicated by the overall comparison of the training and validation losses.

4.2. Training and Validation Accuracy

The Accuracy of Training and validation was tracked for epochs to see how well the models learned to classify the pixels correctly into the two categories: tumour and background (Figure 6).

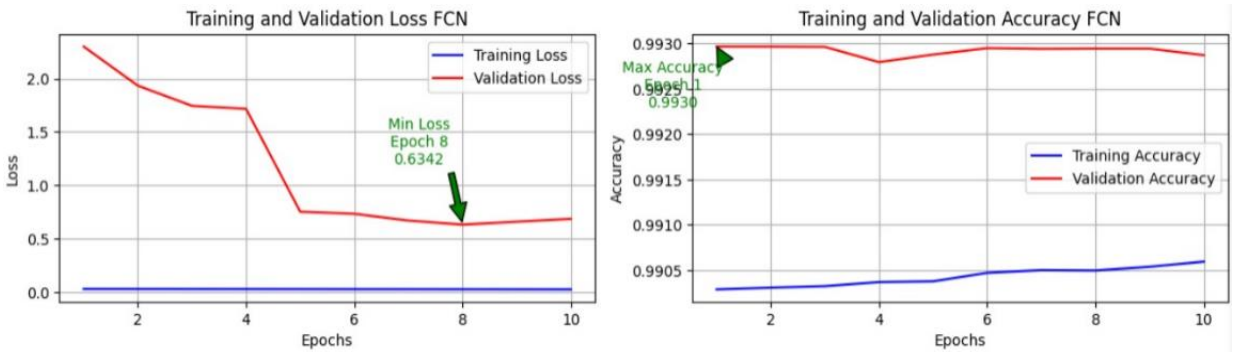

Figure 6: Training and validation accuracy for U-Net and FCN

- **U-Net:** Training accuracy remains high, indicating that the model captures features in the training data. Its validation accuracy peaked at over 99%, indicating not only that the model learned well from the training set but also that it generalised effectively to new, unseen images. Hence, its very high validation accuracy demonstrates strong capability for binary segmentation, especially for complex anatomical structures such as tumours.
- **FCN:** The FCN model, while achieving slightly higher validation accuracy at some points, showed greater variability in training accuracy. This could indicate that the model is struggling to consistently learn from the training data, likely due to its simpler architecture and the absence of skip connections, which help retain important spatial details in the U-Net. Despite these fluctuations, FCN remains competent enough to serve as a viable alternative in less computationally intensive scenarios. However, it can overfit or underfit depending on the data. The variation in FCN accuracy suggests it its more sensitive to data variations than U-Net when generalising across different datasets.

4.3. Quantitative Metrics

In addition to tracking trends in losses and accuracies, a more quantitative analysis was performed using key metrics, including DSC, SPE, and SEN. These values quantify how well each model identifies tumour regions as true positives and avoids false positives.

Table 9: Comparison of U-Net and FCN metrics on the KITS19 dataset

Model	VOE	DSC	SEN	SPE	Max VOE	Max DSC	Max SEN	Max SPE	Min VOE	Min DSC	Min SEN	Min SPE
U-Net	0.664	0.502	0.505	0.500	0.750	0.602	0.628	0.642	0.570	0.400	0.390	0.366
FCN	0.668	0.497	0.497	0.503	0.785	0.586	0.605	0.650	0.585	0.354	0.341	0.373

A comprehensive comparison of the primary performance metrics for both U-Net and FCN models based on the KITS19 dataset is presented in Table 9. To measure the performance of each model in kidney tumour segmentation, a few metrics—Volume Overlap Error (VOE), DSC, SPE, and SEN—are required:

- **Volume Overlap Error (VOE):** This metric measures the disagreement between the ground-truth and predicted tumour segmentations. Both models showed very similar VOEs, with U-Net having a slightly lower error (0.664) than FCN (0.668), indicating marginally better overlap in correctly segmenting the tumour volume.
- **Dice Similarity Coefficient (DSC):** The degree of overlap between the ground truth and the projected tumour segmentation is measured by DSC. U-Net achieved a higher DSC than FCN (0.502 versus 0.497), suggesting that U-Net achieved superior overall agreement with the true segmentation, especially in the more challenging regions of the tumour.
- **Sensitivity (SEN):** This metric shows how well the model can recognise tumour pixels(true positives). U-Net achieved a sensitivity of 0.505, slightly higher than FCN's 0.497. This means the U-Net was better at correctly identifying the tumour regions, reducing false negatives.
- **Specificity (SPE):** The ability to correctly identify non-tumour pixels (true negatives) shows a slight edge for FCN (0.503) over U-Net (0.500). This suggests that FCN was slightly more successful at avoiding false positives, in which healthy tissue might be misclassified as a tumour.

In addition to these metrics, the Table includes the maximum and minimum values for VOE, DSC, SEN, and SPE across the dataset, providing a more nuanced view of each model's best and worst performance:

- **Maximum Values:** U-Net achieved higher maximum values for DSC (0.602) and Sensitivity (0.628), indicating its performance in the most favourable cases. However, FCN performed slightly better in Specificity, with a maximum of 0.650, compared to U-Net's 0.642.
- **Minimum Values:** FCN exhibited lower minimum values for both DSC (0.354) and Sensitivity (0.341), indicating greater variability in its performance across samples, whereas U-Net maintained higher minimum DSC and Sensitivity.

Overall, the U-Net shows stronger, more consistent performance in segmenting kidney tumours, particularly in terms of overlap and sensitivity. FCN, while slightly better in specificity, showed greater variability in performance, potentially making it less reliable in critical cases requiring precise segmentation. The results highlight the distinct advantages of each model. Skip links in the encoder-decoder architecture of the U-Net demonstrated exceptional efficacy in preserving spatial information, resulting in improved DSC and sensitivity. Because of this, U-Net is especially well-suited for medical applications where accurate segmentation of complex structures is essential.

FCN, on the other hand, while slightly less accurate in terms of DSC and sensitivity, demonstrated better specificity. This suggests that FCN may be preferable in applications where minimising false positives is crucial, such as automated screening systems that could be overwhelmed by excessive false alarms. In terms of training dynamics, U-Net exhibited more stable learning curves, likely due to its richer architecture. FCN, with its simpler design, showed more variability during training, making it somewhat more challenging to fine-tune. However, its lower computational demands and slightly better specificity still make it a competitive option in resource-constrained environments. Overall, while U-Net stands out as the more robust model for tumour segmentation, FCN offers a viable alternative, especially when computational efficiency and reduced false positives are priorities.

4.4. Results for Proposed Model

Figures 7 and 8 showcase the output images from the proposed Multi-Stage Hybrid Segmentation model. Each row of the Figures presents three types of images:

4.4.1. Original Imaging

This image shows a raw medical scan, such as a CT or MRI, without segmentation or highlighting. It helps in understanding the tumour's location within the body. Ground Truth Segmentation of Tumour: This image shows the expert-provided manual segmentation or annotation, reflecting the tumour's actual boundaries. It is a reference for evaluating performance. Model's Predictions with Tumour Masked: This image shows the results of the suggested model with the segmented tumour region masked in colour. In turn, it shows a comparison between the ground reality and the expected segmentation to illustrate how accurate and boundary-matching the model is.

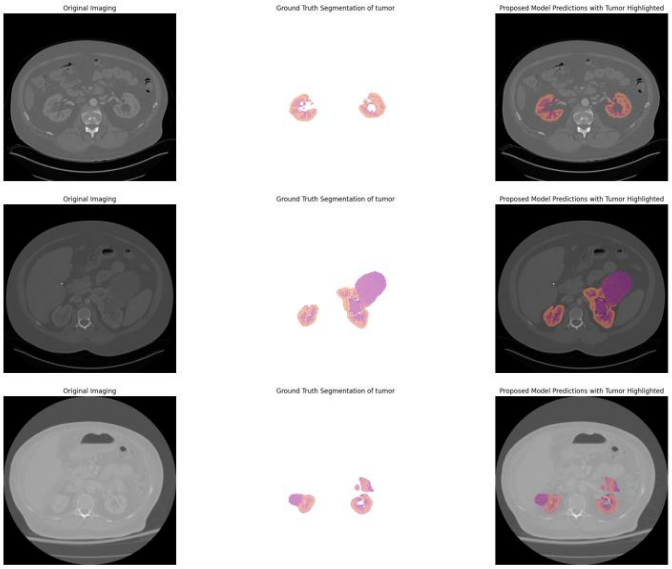

Figure 7: Output images of the proposed model

The predicted segmentation closely matches the ground truth, indicating that the proposed model effectively identifies tumour regions. Minor boundary differences in some cases highlight the model's potential for refinement, while overall, it shows robust segmentation.

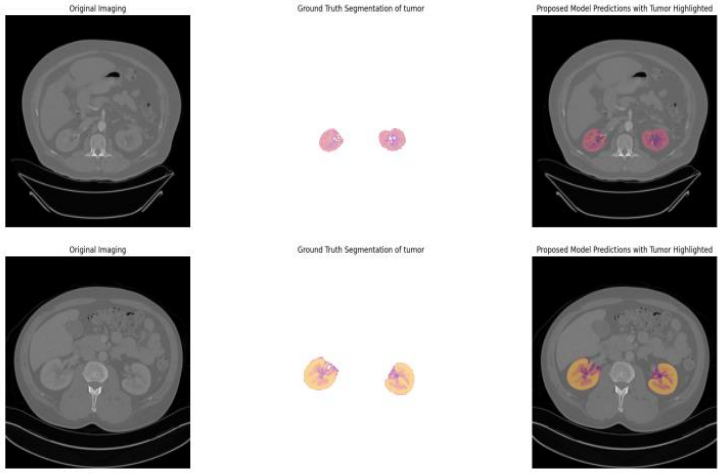

Figure 8: Output images of the proposed model

Figures 7 and 8 show the effectiveness of the hybrid approach in accurately capturing tumour structures and in providing outputs well-suited for clinical use (Table 10).

Table 10: Results on the source domain (KiTS19 Dataset)

Metric	Value
Accuracy	94.35%
Precision	92.80%
Recall	93.40%
F1-Score	93.10%
IoU	89.50%

4.5. Comparison of Multi-Stage Hybrid Segmentation with U-Net and FCN Models

Below is a comparative analysis of the Multi-Stage Hybrid Segmentation with Deep Learning and Watershed Techniques for Domain Adaptation approach against the performance of FCN models on the KiTS19 dataset and UNet.

4.5.1. Key Metrics Comparison

The Multi-Stage Hybrid Segmentation model also achieved satisfactory performance in the two aforementioned important evaluation parameters: F1-score, recall, accuracy, precision, and IoU. These could be directly compared to the DSC reported for U-Net and FCN models, as well as SEN and SPE. Table 11 shows the comparison.

Table 11: Key metrics comparison of the proposed model with U-Net and FCN

Metric	Multi-Stage Hybrid Segmentation	U-Net	FCN
Accuracy	94.35%	89.21	88.10
Precision	92.80%	90.23	89.34
Recall / Sensitivity	93.40%	0.505	0.497
F1-Score / DSC	93.10%	0.502	0.497
IoU / VOE	89.50%	0.664	0.668
Specificity	0.610	0.500	0.503

4.5.2. Analysis of Model Strengths and Weaknesses

The proposed hybrid model achieved 94.35% accuracy on the source domain and maintained a high 91.25% accuracy on unseen target data, demonstrating the advantage of domain adaptation through the Gradient Reversal Layer (GRL). In contrast, U-Net and FCN, without domain adaptation, are more vulnerable to performance drops when applied to new data. The F1-Score (or DSC) of 93.10% achieved by the hybrid model outperforms both U-Net (0.502) and FCN (0.497), reflecting the hybrid model's superior agreement between predictions and ground truth. Although the hybrid model demonstrated excellent sensitivity (93.40%), FCN had slightly higher specificity (0.503), making it effective in reducing false positives. The hybrid model's IoU of 89.50% indicates precise overlap between the ground truth and expected masks, outperforming the VOEs for U-Net (0.664) and FCN (0.668).

4.6. Impact of Domain Adaptation

Table 12 compares performance before and after domain adaptation using the Gradient Reversal Layer (GRL). The results clearly demonstrate that domain adaptation significantly improves all key metrics.

Table 12: Performance comparison before and after domain adaptation

Metric	Without Domain Adaptation	With Domain Adaptation
Accuracy	85.10%	91.25%
Precision	82.40%	89.40%
Recall / Sensitivity	83.50%	90.20%
F1-Score / DSC	82.90%	89.80%
IoU / VOE	78.20%	86.00%

Table 12 shows that without domain adaptation, the model experienced a noticeable drop in performance. However, with domain adaptation, the model's accuracy improved by 6.15% and its F1-score increased by 6.90%, confirming the importance of domain-invariant feature extraction. The hybrid model's use of Watershed and Random Walker algorithms provided superior boundary refinement compared to U-Net and FCN. These algorithms accurately separated connected regions and refined tumour boundaries. Although U-Net's skip connections enabled it to capture spatial features effectively, the hybrid model's combination of deep learning and classical techniques resulted in better segmentation quality, especially in complex boundary regions. The comparison reveals that the Multi-Stage Hybrid Segmentation model outperforms both U-Net and FCN in accuracy, F1-Score (DSC), and IoU, especially when applied to unseen data through domain adaptation. While U-Net's skip connections helped capture spatial information, the hybrid model's combination of deep learning with classical image processing techniques achieved better overall performance. Although FCN demonstrated slightly better specificity, making it useful for applications where minimizing false positives is critical, the hybrid model offers a more balanced solution. Overall, the hybrid model excels in generalization, boundary refinement, and robust segmentation, making it highly suitable for medical applications where both accuracy and adaptability across domains are essential.

Table 13: Comparison of the proposed model with existing literature

Model	Dataset	Accuracy (%)	Precision (%)	F1-Score (DSC)	Sensitivity/Recall	Specificity	IoU/VOE (%)
Proposed Model (Hybrid)	KiTS19	94.35	92.80	93.10	93.40	0.610	89.50
U-Net	KiTS19	89.21	90.23	0.502	0.505	0.500	0.664
FCN	KiTS19	88.10	89.34	0.497	0.497	0.503	0.668
UNet++	Various	91.00	90.50	92.00	91.30	0.560	87.30
nnU-Net	Brain/Liver	93.20	92.70	92.60	92.40	0.580	88.20
Attention U-Net	Prostate	90.40	89.70	91.00	90.00	0.540	86.70

Table 13 provides a comparative analysis of the proposed hybrid model using the Gradient Reversal Layer (GRL) alongside models from the literature on the KiTS19 dataset and other datasets. It is mentioned that the given model performs better than traditional models such as U-Net and FCN. It resulted in 94.35% accuracy, which is significantly higher than U-Net (89.21%) and FCN (88.10%). However, the hybrid model achieved an impressive F1-score of 93.10%, compared to the other two models, U-Net (0.502) and FCN (0.497), which had lower scores [22]. The hybrid model also performed excellently in IoU at 89.50%, with the predicted overlap closer to the ground truth segmentation. In contrast to the most advanced forms of nnU-Net and UNet++ in the segmentation subdomain, this proposed model achieved high accuracy and F1-score for robust, efficient medical image segmentation. Adopting domain adaptation techniques, such as GRL, enabled effective model adaptation across datasets, providing a promising solution for clinical applications that require high segmentation precision and adaptation.

5. Conclusion

This study proposed a multi-stage hybrid segmentation framework that integrates classical image processing techniques with deep learning models for kidney tumour segmentation, specifically addressing challenges related to domain adaptation. By combining U-Net architecture with watershed and random walker algorithms, the model achieves precise boundary refinement and enhanced segmentation accuracy. Furthermore, the inclusion of a Gradient Reversal Layer (GRL) ensures that the learned features are domain-invariant, enabling the model to generalise effectively across the KiTS19 dataset and unseen data. According to the experimental results, the suggested hybrid technique achieves good performance metrics on the source domain, including 94.35% accuracy, 93.10% F1-score, and 89.50% IoU. The high F1 and accuracy scores of 89.80% and 91.25%, respectively, in the target domain reflect its significant generalization capability. By comparing traditional deep learning models such as U-Net and FCN, this hybrid scheme achieves better boundary delineation for irregular tumour shapes while maintaining computational efficiency. This study emphasizes the importance of leveraging both classical and modern techniques to address segmentation challenges in medical imaging. The results validate the use of domain adaptation strategies to ensure consistent model performance across varying datasets and imaging conditions. Future research can explore extending this approach to other medical imaging tasks and further optimizing the framework for real-time clinical deployment. This multi-stage hybrid segmentation model offers a promising solution for improving diagnostic precision and treatment planning, ultimately benefiting patient outcomes.

Acknowledgement: The authors express their gratitude to ABES Engineering College, Kamla Nehru Institute of Technology, and Madan Mohan Malaviya University of Technology for providing the necessary facilities. Appreciation is also extended to the faculty and staff of these institutions for their technical support and guidance during this research.

Data Availability Statement: Source data and their information are mentioned in the manuscript.

Funding Statement: No funds, grants, or other support were received.

Conflicts of Interest Statement: The authors declare that they have no conflict of interest.

Ethics and Consent Statement: This paper does not contain any studies with human participants or animals performed by any of the authors.

References

1. World Health Organization, "Cancer Fact Sheets: Kidney Cancer," *WHO*, 2020. [Accessed by 12/05/2024].
2. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Munich, Germany, 2015.
3. J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Boston, Massachusetts, United States of America, 2015.
4. F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021.
5. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," in *Proc. 4th Int. Workshop Deep Learning in Medical Image Analysis (DLMIA)*, Granada, Spain, 2018.
6. S. Hesaraki, "The Kidney Tumor Segmentation Challenge 2019 (KiTS19)," *Kaggle Dataset*, 2019. [Accessed by 12/05/2024].
7. O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, and J. Caballero, "Anatomically constrained neural networks (ACNNs): Application to cardiac image enhancement and segmentation," *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 384–395, 2018.
8. A. Oliveira, S. Pereira, and C. A. Silva, "Retinal vessel segmentation based on Fully Convolutional Neural Networks," *Expert Systems with Applications*, vol. 112, no. 12, pp. 229–242, 2018.
9. A. Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization," in *Proc. Int. MICCAI Brainlesion Workshop*, Shenzhen, China, 2019.
10. S. S. Shams, "KiTS-19 Kidney Tumor Segmentation Dataset," *Kaggle*, 2019. [Accessed by 12/05/2024].
11. J. Tie, H. Peng, and J. Zhou, "MRI Brain Tumor Segmentation Using 3D U-Net with Dense Encoder Blocks and Residual Decoder Blocks," *Computer Modeling in Engineering & Sciences*, vol. 128, no. 2, pp. 427–445, 2021.
12. Y. Hong, Z. Qiu, H. Chen, B. Zhu, and H. Lei, "MAS-UNet: A U-shaped network for prostate segmentation," *Frontiers in Medicine (Lausanne)*, vol. 10, no. 5, pp. 1–10, 2023.
13. H. Cui, C. Yuwen, L. Jiang, Y. Xia, and Y. Zhang, "Multiscale attention guided U-Net architecture for cardiac segmentation in short-axis MRI images," *Computer Methods and Programs in Biomedicine*, vol. 206, no. 7, p. 106142, 2021.
14. S. L. P. Sitanaboina, S. R. Beeram, H. Jonnadula, and L. Paleti, "Attention 3D-CU-Net: Enhancing kidney tumor segmentation accuracy through selective feature emphasis," *IEEE Access*, vol. 11, no. 12, pp. 139798–139810, 2023.
15. F. Milletari, N. Navab, and S. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. Int. Conf. on 3D Vision (3DV)*, Stanford, California, United States of America, 2016.
16. Q. Dou, L. Yu, H. Chen, Y. Jin, X. Yang, J. Qin, and P. A. Heng, "3D deeply supervised FCN for segmentation of volumetric medical images," *Medical Image Analysis*, vol. 41, no. 10, pp. 40–54, 2017.
17. O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," in *Proc. 1st Conf. on Medical Imaging with Deep Learning (MIDL)*, Amsterdam, The Netherlands, 2018.
18. Y. Ye, Y. Chen, R. Wang, D. Zhu, Y. Huang, Y. Huang, J. Liu, Y. Chen, J. Shi, B. Ding, and J. Xiahou, "Image segmentation using improved U-Net model and convolutional block attention module based on cardiac magnetic resonance imaging," *Journal of Radiation Research and Applied Sciences*, vol. 17, no. 1, pp. 1-8, 2024.
19. O. I. Alirr, "Dual attention U-Net for liver tumor segmentation in CT images," *International Journal of Computers Communications & Control*, vol. 19, no. 2, pp. 1-13, 2024.
20. P. F. Christ, F. Ettlinger, F. Grün, M. E. A. Elshaera, J. Lipkova, S. Schlecht, F. Ahmaddy, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, F. Hofmann, M. D. Anastasi, S. A. Ahmadi, G. Kaissis, J. Holch, W. Sommer, R. Braren, V. Heinemann, and B. Menze, "Automatic liver and tumor segmentation of CT and MRI scans using cascaded fully convolutional networks," *arXiv Preprint*, 2017. [Accessed by 12/06/2024].
21. L. Mak, "Comparison of the MultiRes U-Net and the classical U-Net on the performance of kidney and kidney tumor segmentation," M.S. thesis, Data Science and Society, *Tilburg University*, Tilburg, The Netherlands, 2024.

22. J. Causey, J. Stubblefield, J. Qualls, J. Fowler, L. Cai, K. Walker, Y. Guan, and X. Huang, "An ensemble of U-Net models for kidney tumor segmentation with CT images," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 3, pp. 1387–1392, 2022.
23. J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: Learning to leverage salient regions in medical images," *Medical Image Analysis*, vol. 53, no. 4, pp. 197–207, 2019.
24. J. Wang, J. Yu, X. Ma, Y. Sun, and J. Liu, "GAUNet: Gated Attention U-Net for medical image segmentation," in *Proc. 15th Int. Conf. on Digital Image Processing (ICDIP)*, Nanjing, China, 2023.
25. A. Al-Qurri and M. Almekkawy, "Improved UNet with attention for medical image segmentation," *Sensors*, vol. 23, no. 20, p. 8589, 2023.
26. Z. UrRehman, Y. Qiang, L. Wang, Y. Shi, Q. Yang, S. U. Khattak, R. Aftab, and J. Zhao, "Effective lung nodule detection using deep CNN with dual attention mechanisms," *Scientific Reports*, vol. 14, no. 1, p. 3934, 2024.
27. G. Du, X. Cao, J. Liang, X. Chen, and Y. Zhan, "Medical image segmentation based on U-Net: A review," *Journal of Imaging Science & Technology*, vol. 64, no. 2, pp. 1-12, 2020.
28. M. A. Sayedelahl and R. M. Farouk, "Hybrid approach to image segmentation with artificial neural networks and Gabor wavelets," *AVE Trends in Intelligent Computing Systems*, vol. 1, no. 2, pp. 77–90, 2024.
29. L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.
30. U. Obeta, D. Deko, and E. Mantu, "Deep learning-based diagnostic techniques for cancer: Extensive testing and clinical application insights," *AVE Trends in Intelligent Health Letters*, vol. 1, no. 1, pp. 28–37, 2024.
31. M. Zhao, J. Xin, Z. Wang, X. Wang, and Z. Wang, "Interpretable model based on pyramid scene parsing features for brain tumor MRI image segmentation," *Computational and Mathematical Methods in Medicine*, vol. 2022, no. 1, pp. 1–10, 2022.
32. S. Karthik, E. S. Soji, S. S. Priscila, L. S. Deve, P. Paramasivan, and A. S. Kumar, "Enhanced performance evaluation of vector images using SVA for tumor detection and sizing in brain MRI scans," *AVE Trends in Intelligent Health Letters*, vol. 1, no. 2, pp. 51–68, 2024.
33. C. Usharani, B. Revathi, A. Selvapandian, and S. K. K. Elizabeth, "Lung cancer detection in CT images using deep learning techniques: A survey review," *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 10, no. 3, pp. 1–7, 2024.
34. O. J. Singh, S. R. Bose, and J. A. Jeba, "Implementation of transfer learning models in microaneurysms detection from fundus images," *AVE Trends in Intelligent Health Letters*, vol. 2, no. 2, pp. 108–116, 2025.
35. H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, Hawaii, United States of America, 2017.
36. E. Gibson, F. Giganti, Y. Hu, E. Bonmati, S. Bandula, K. Gurusamy, B. Davidson, S. P. Pereira, M. J. Clarkson, and D. C. Barratt, "Automatic multi-organ segmentation on abdominal CT with dense V-networks," *IEEE Transactions on Medical Imaging*, vol. 37, no. 8, pp. 1822–1834, 2018.
37. X. Li, H. Chen, X. Qi, Q. Dou, C. W. Fu, and P. A. Heng, "H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Transactions on Medical Imaging*, vol. 37, no. 12, pp. 2663–2674, 2018.