**Control and Cybernetics**
vol. **54** (2025) No. 1
pages: 79-117
DOI: 10.2478/candc-2025-0004

# Insights into U-NET models with special focus on ultrasound and MRI medical image segmentation*

by

**V. B. Shereena and G. Raju**

MES College Nedumkandam, Chembalam PO, Idukki District,
Kerala-685 553, India

**Abstract:** The advent of deep learning enabled the extraction of complex feature representations from medical imaging data, which was considered impossible to be achieved with standard computer learning. The applications of deep learning in the field of medical image analysis afford significant results. A key feature of deep learning techniques is their ability to automatically learn task-specific feature representations and extract relevant features without human intervention. Various deep learning models, including CNN, AlexNet, ResNet, DenseNet and U-Net were developed for medical image analysis. Among these models, U-Net is a popular model, used for medical image segmentation. The present article provides a comprehensive review of the deep learning segmentation models, which use U-Net and its variants, applied in the domain of medical image segmentation, specifically tailored to medical imaging modalities, such as ultrasound and MRI, along with respective pros and cons in the field of image segmentation. The analysis reveals that the performance of different U-Net variants varies significantly based on imaging modality and segmentation complexity.

**Keywords:** medical imaging, ultrasound, MRI, segmentation, deep learning, U-Net

## 1. Introduction

### 1.1. Image segmentation problem and techniques

Medical imaging has been for a long time and very widely used as an essential and effective technique for clinical diagnosis. Among the medical imaging modalities, X-rays, Magnetic Resonance Imaging, Ultrasound and Positron

*Submitted: August 2024; Accepted August 2025.

Emission Tomography are frequently used in the early-stage diagnosis for clinical decision making, see Ahmad, Yu and Miller (2013). Radiologists diagnose disorders by analysing the images acquired by these techniques. MRI systems offer the advantage of avoiding ionizing radiation unlike X-ray imaging and are capable of acquiring images in any plane effectively. Ultrasound imaging is a portable, cost-effective modality that does not involve the use of ionizing radiation, neither. Segmentation plays a crucial role in finding the critical information from medical images (Hesamian et al., 2019). Numerous studies have been conducted in the field of medical image segmentation for getting faster, accurate, and better results. However, designing efficient and robust algorithms for precise and correct segmentation of medical images remains a significant challenge. In this article, we present a comprehensive review on various deep learning approaches, based on U-Net, and its variants, focusing on their application in the segmentation of ultrasound and MRI images.

In the context of medical imaging, segmentation refers to the process of dividing an image into non-overlapping regions of pixels that correspond to significant anatomical structures or abnormalities, such as tumors or cysts, see Kumar et al. (2018a). Accurate segmentation of organ images is crucial for various diagnostic and surgical procedures, and it also supports applications like computer-aided diagnosis and image-guided surgery. Traditional segmentation algorithms are categorized into first, second and third generation algorithms, conform to the evolution of techniques in terms of automation, adaptability, and reliance on human input, Kumar et al. (2018a). First generation algorithms include thresholding approaches, see, e.g. Houssein et al. (2021), edge based approaches and region based approaches, which rely heavily on hand-crafted features and prior knowledge, see Priyadharsini and Sharmila (2019). Second generation algorithms include the methods of active contours, Yu, He and Pan (2019), clustering, Ezugwu et al. (2020), watershed, and Markov random fields methods, Sridhar (2020), where features were still manually designed, but classifiers learned to map them to output labels. Third generation algorithms employ artificial neural networks, graph theory and related techniques, as well as hybrid approaches and deep learning techniques, which automate both feature extraction and classification, Kumar et al. (2018a). U-Net and its variants, explored in this article, are representative of the third-generation segmentation approaches.

The advent of third generation segmentation algorithms, using deep learning approaches outweighed various traditional methods, adopted for segmentation of medical images in the field, see Altaf et al. (2019). Deep learning is associated with the family of artificial neural networks (ANN) algorithms, where hyper parameters, including learning rate and number of epochs, govern the network structure. Various deep learning models, including Convolutional Neural Networks (CNNs), AlexNet, ZFNet, VGG, GoogleNet, ResNet, and DenseNet were originally developed, in particular – for general image classification tasks, Zhou, Ruan and Canu (2020). In contrast, U-Net was specifically designed for biomedical image segmentation and has become a foundational architecture in this field. U-Net, having evolved from the traditional convolutional neural network, was first designed and applied in 2015 for the segmentation of light microscopy biomedical images, Du et al. (2020). U-Net consists of a contracting path and an expansive path, this giving it the U-shaped architecture. It produces accurate results with appropriate training datasets, and with a carefully selected number of training epochs to avoid underfitting or overfitting. The limitations of U-Net include high GPU memory requirements for processing high-resolution images, longer computational times, and the lack of pre-trained models, due to its application-specific nature.

Various researchers have published different variants of U-Net, and this article discusses the most popular models among them. An extension of the basic U-Net framework, 3D U-Net enables effective 3D volumetric segmentation. This model has been extensively used in volumetric Computer Tomography (CT) and MR image segmentation applications, see Siddique et al. (2021). A variant of U-Net, U-Net++, with skip pathways, performed more effective segmentation of medical images than the original U-Net, see Zhou et al. (2018). Another modified version of U-Net, Attention U-Net, provides localized classification information, while ignoring unnecessary areas, Siddique et al. (2021), so as to effectively classify different objects. Recurrent U-Net, an extended version of U-Net, adds a recurrent unit over some of its encoder and decoder paths and operates efficiently even in the environments with limited training data and computational power, Wang et al. (2019). Residual U-Net architecture solves the vanishing gradient problem in deeper neural networks by utilizing skip connections, Siddique et al. (2021). Yet another variant of U-Net, named Dense U-Net, integrates two advanced components into the U-Net architecture: an Inception-Res module, which combines the multi-scale feature extraction of Inception blocks with residual connections to improve gradient flow; and a densely connected convolutional module, inspired by DenseNet, where each layer receives inputs from all preceding layers to enhance feature reuse and reduce the number of parameters. This architecture showed better performance on CT and MRI datasets than the basic U-Net model (Zhang et al., 2020). Then, R2U++, an advanced deep learning architecture that builds upon the Recurrent Residual U-Net (R2U-Net) and U-Net++ frameworks, overcomes the issues due to complicated datasets in medical image segmentation architecture by using a deeper recurrent residual convolution block to extract crucial features and dense skip pathways in order to accumulate features coming from multiple layers and apply concatenation accordingly, Mubashar et al. (2022).

### 1.2. Motivation and contribution

U-Net models, mentioned above, played a major role in the analysis of medical images obtained from various modalities of imaging, namely MRI, magnetic resonance imaging and USG, ultrasound medical image segmentation in the recent years. Many researchers have published reviews on medical image segmentation. Understanding the available methods is essential for the development of computer aided diagnosis systems. This motivated us to perform a literature review on variants of U-Net and summarise the relevant methods, adopted to segment and analyse the MRI and USG medical images with the assumption of application of U-Net models. This article introduces such a review to accentuate the development in the field of U-Net architecture in diagnosing the medical images.

The major contributions of this research work are as follows:

- Provision of an extensive survey of the network architecture of the key U-Net based architectures used in medical image segmentation, outlining their design differences and evaluating their strengths and limitations across specific imaging modalities such as MRI and USG.
- Discussion of the results and an outline for the future scope for the aforesaid research area to aid researchers in making further contributions in terms of investigations in this area.

The remainder of the article is structured as follows. Section 2 contains the detailed content review of studies, devoted to U-Net models; Section 3 explains the architecture of various U-Net models; Section 4 presents the application of U-Net models in the segmentation of ultrasound and MRI images of liver, lung, prostate, pancreas and brain, respectively; Section 5 presents the performance metrics used for evaluating segmentation models and Section 6 presents results and discussion of application of U-Net models in image analysis; finally, Section 7 presents the conclusions reached from the review.

## 2. Related works

This section provides a presentation of a broad selection of a focussed literature of U-Net architecture and its popular versions.

Ronneberger, Fischer and Brox (2015) proposed the basic U-Net model for the segmentation of microscopic images, which outperformed the standard convolutional neural network architecture. The proposed U-Net architecture contains a U-shaped contracting path, named encoder, and an expanding path, named decoder. The dataset consisted of 30 microscopy images, acquired from the ISBI 2012 challenge. This architecture has the merit of faster segmentation using GPU, Graphical Processing Unit systems, with the use of a limited dataset of 30 microscopic images.

Initially, 2D U-Net segmentation methods were used for the segmentation of medical images. But the possibility of misclassification of images is a prominent challenge in the 2D detection of brain tumors. Hence, automatic segmentation of 3D MRI images of the brain can greatly improve diagnosis and treatment procedures, see Sangui et al. (2023). The proposed 3D model has been evaluated using MRI images of BRATS 2020 dataset and achieved test accuracy of 99.4%, outperforming VGG16 and ResNet50, which yielded the accuracies of 98.7% and 97.2%, respectively.

The basic U-Net has two main limitations. First, it requires an extensive search to find the best number of layers (also called network depth) for a specific task. Second, its skip connections can only combine features from layers at the same scale or resolution. This makes it challenging for the network to learn from both fine details and broader context at the same time. To overcome these limitations, a powerful variant of U-Net named U-Net++ was developed by Zhou et al. (2020) for medical image segmentation. In this model, nested and dense skip connections were added to encoder and decoder sub-networks, in order to reduce the semantic gap between the feature maps of the encoder and decoder sub-networks. This makes learning an easier task. Experiments were conducted to evaluate U-Net++ using six different medical image segmentation datasets, including CT, MRI and electron microscopy images. The results of comparison between the proposed model and other U-Net architectures demonstrated that U-Net++ consistently outperformed the base U-Net models across various datasets, enhancing the segmentation quality of objects with varying sizes.

The Attention U-Net architecture improves the network performance over U-Net++, but is sensitive to the datasets. Chen et al. (2023b) developed an improved attention U-Net, a hybrid adaptive attention module for the segmentation of breast ultrasound images. The model consisted of a channel self-attention block and a spatial self-attention block, instead of the convolution operation. Experiments on three public breast ultrasound datasets demonstrated that this method has an improved performance on breast ultrasound segmentation with better generalization performance and flexibility than the existing network frameworks. While the method shows promising results, like most approaches, it requires further optimization to handle the complexity of breast ultrasound data and to obtain accurate contours.

Wang et al. (2019) introduced a recurrent U-Net architecture for the segmentation of medical images. The model integrates recursions on the predicted segmentation mask and multiple internal states of the network. This model is intended to have efficient performance with limited resources. The model was experimentally verified for hand, retina vessel, road, and urban landscape segmentations and demonstrated its outstanding performance compared to the then recent methods.

Another performance enhancement in the U-Net model was attention residual U-Net (AResU-Net), introduced by Zhang et al. (2020), which integrated attention mechanism and residual units into U-Net for brain tumor segmentation. AResU-Net was extensively evaluated on BraTS 2017 and BraTS 2018 MRI brain tumor segmentation datasets. Experimental results demonstrated that the proposed model attains comparable performance with typical brain tumor segmentation methods. The model has the limitation of utilizing only 2D images due to resource constraints, which resulted in partial loss of contextual information.

Wang et al. (2019) proposed a dense U-Net model, which eliminates the vanishing gradient problem and improves network performance compared to Residual U-Net. The Dense U-Net contains three densely connected blocks, selective de-convolution layers with up-sample and trans-convolution filling function. Dense U-Net was evaluated using liver and spleen CT dataset and the comparison of results with other state of the art methods proved its robustness and efficiency with limited data. The accuracy of the model was improved with a customized loss function so as to get an average Dice ratio for the liver and spleen data at 94.9% and 92.1%, respectively which was higher than that of other models including FCN (liver: 92.1%, spleen: 88.9%) and classical U-Net (liver: 92.6%, spleen: 89.3%).

Like with the most of deep learning-based segmentation methods, the performance of U-Net variants can be affected by the challenges such as noise, variability in anatomy, and low contrast in complex medical imaging datasets. Mubashar et al. (2022) proposed a new U-Net based model, R2U++, to overcome these limitations. In this model, the convolutional block is replaced by a deeper recurrent residual convolution block to extract crucial features for segmentation. A dense skip pathway is introduced to reduce the semantic gap between encoder and decoder, which integrates features coming from multiple scales and applies concatenation. Extensive experiments were conducted on four distinct medical imaging modality datasets: ISBI 2012 electron microscopy images, COVID-19 CT images, JSRT chest X-ray dataset, and the DRIVE fundus image dataset. The results demonstrated improved performance, with an average gain in IoU of $1.5 \pm 0.37\%$ and in Dice score of $0.9 \pm 0.33\%$ over U-Net++, and gains of $4.21 \pm 2.72\%$ in IoU and $3.47 \pm 1.89\%$ in Dice score over R2U-Net across the different datasets.

An automated and efficient method is necessary for breast ultrasound images classification to assist experts in the diagnosis of tumors. Inspired by the U-Net model and its many variants, Chen et al. (2023a) proposed a simple nested U-

Net (NU-Net) for accurate and efficient breast tumor segmentation. Extensive experimental results on three public breast ultrasound datasets revealed that this method has a better segmentation performance on breast tumors compared to fifteen state of the art segmentation methods, including widely used U-Net variants like U-Net++, Attention U-Net, multi-scale U-Nets, and other complex architectures.

Although U-Net and its variants have achieved remarkable success in medical image segmentation, one key limitation is the semantic gap between encoder and decoder feature maps. Encoder features preserve fine-grained spatial details but lack strong semantic representation, whereas decoder features capture high-level semantics but lose spatial precision. Reducing the semantic gap is essential for improving feature fusion between encoder and decoder, thereby enhancing segmentation efficiency and accuracy.

## 3. The U-Net models

This section describes the architecture of the basic U-Net model and then of the different variants of the basic U-Net model, which are quite commonly used for medical image segmentation.

### 3.1. Basic U-Net

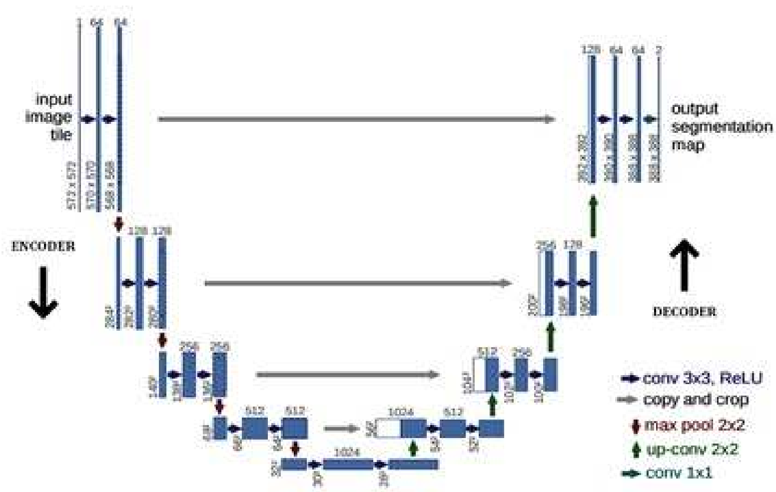


Figure 1. The basic U-Net model, after Ronneberger, Fischer and Brox (2015)

The basic U-Net architecture consist of an encoder and decoder of depth 4 with an input image size of 256 x 256, Ronneberger, Fischer and Brox (2015). The encoder consists of two successive 3 x 3 convolution layers, followed by a rectified linear unit (ReLU) activation layer, then a $2 \times 2$ max-pooling layer with stride 2 for down-sampling. In the decoder, up-sampling is achieved by 2 x 2 up-convolution layer and concatenation to the up sampled feature map in each depth. In the final stage, 1 x 1 convolution layer with a sigmoid activation function is applied to reduce the feature map to the required level and produce the segmented image, Siddique (2021). The scheme of the basic U-Net architecture is shown in Fig. 1.

### 3.2. 3D U-Net

3D U-Net extends the U-Net architecture by replacing all 2D operations with corresponding 3D equivalents, Sangui et al. (2023), Cicek et al. (2016). This model introduced volumetric segmentation, where the training dataset is a set of sparsely annotated volumetric images.

Initially, BraTS dataset images are normalised and resized. In the encoder part, the standard convolutions and a max-pooling unit (2 x 2) is used at each level. The depth of the feature maps gradually increases, while the spatial resolution of the image decreases, progressing from 128x128x2 to 8x8x512. In the decoder part, transposed convolutions are used along with standard convolutions to gradually expand the image size, while reducing the depth, from 8x8x512 to 128x128x2. The skip connections in the encoder are combined with the decoder layers, helping U-Nets to generate an image in the decoder using detailed information learned in the encoder. The loss function, which is used, is categorical cross-entropy and the optimizer is Adam. The scheme of the 3D U-Net architecture is shown in Fig. 2.

### 3.3. U-Net++

Another improved version of U-Net is U-Net++, in which nested and dense skip connections were added and each decoder layer got connected to multiple encoder layers, Zhou et al. (2018). The model used a series of nested sub U-Nets with increasing depths in order to efficiently utilize features from multiple abstraction levels. The number of convolution layers in U-Net++ depends on the pyramid level. The skip pathway is formulated as given below:

Let $x_{a,b}$ be the output of the node $X_{a,b}$, where $a$ is the down sampling layer along the encoder, and $b$ is the convolution layer of the dense block along the

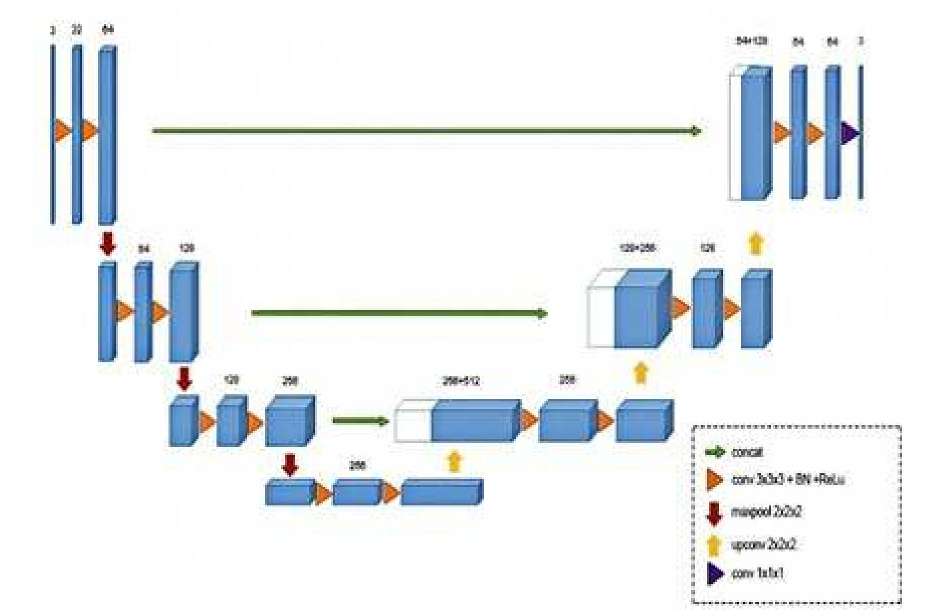


Figure 2. The 3D U-Net model, after Cicek et al. (2016) (in the 3D U-Net, all convolutional, pooling, and upsampling operations are extended into 3D, enabling volumetric context capture)

skip pathway. The stack of feature maps, represented by $x_{a,b}$, is calculated as in equation (1), Zhou et al. (2018):

$$x^{a,b} = \left\{ \begin{array}{l} H\left(x^{a-1,\ b}\right),\ b=0 \\ H\left(\left[x^{a,b}\right]_{c=0}^{a-1}\right),\ U(x^{a+1,\ b-1})),\ b>0 \end{array} \right\}, \tag{1}$$

where function $H$ is a convolution operation, followed by an activation function, [ ] denotes the concatenation layer, and $U$ corresponds to an up-sampling layer. Nodes at depth $b = 0$ receive only one input from the previous layer of the encoder; nodes at level $b = 1$ receive two inputs, both from the encoder sub-network, but at two consecutive levels; and nodes at levels $b > 1$ receive $b + 1$ inputs, of which $b$ inputs are the outputs of the previous $b$ nodes in the same skip pathway and the last input is the up-sampled output from the lower skip pathway. The U-Net++ model is depicted in Fig. 3.

U-Net++ performs better than the basic U-Net by using skip connections to combine low-level features from the encoder with high-level features from the decoder.

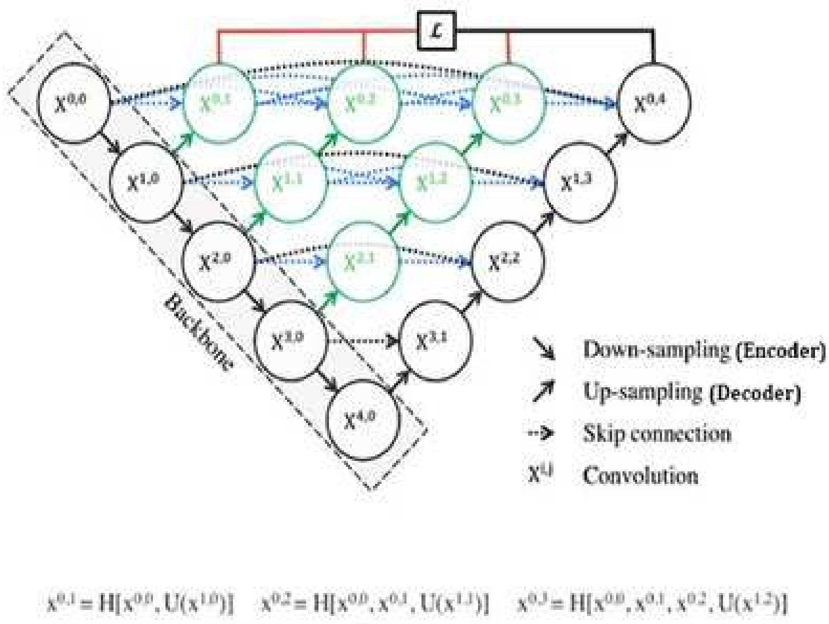


Figure 3. UNet++ model, after Zhou et al. (2018)

### 3.4. Attention U-Net

A modified version of U-Net, named Attention U-Net, introduced attention mechanisms to emphasize important features while ignoring unnecessary areas during segmentation, see Oktay et al. (2018). The attention U-Net includes an attention gate module that evaluates various feature maps according to importance, enabling the model to concentrate on pertinent areas and thereby to increase accuracy. Multiscale attention enabled U-Net model (MA-UNet) can also be efficiently used for segmentation of relevant features from medical images at different scales. It has two branches; first branch acts as a multi-scale feature extraction block by using dense connections and the second branch uses the channel and spatial attention network to suppress irrelevant information, see Punn and Agarwal (2022). Figure 4 shows the schematic structure of the attention U-Net model.

The corresponding features from the contracting path must pass through an attention gate in each layer of the expansive path before the features are concatenated with the up sampled features in the expansive path, Chen et al.

(2023b). The segmentation performance is significantly enhanced by repeatedly using the attention gate after each layer, without considerably increasing the model's computational complexity. Additive attention is formulated as follows, see Oktay et al. (2018):

$$q_{att}^{l} = \psi^{T}\ \left(\sigma_1\left(W_x^T\, x_i^l + W_g^T\, g_i + b_g\right)\right) + b_\Psi \tag{2}$$

$$\alpha_i^l = \sigma_2(q_{att}^l(x_i^l g_i;\, \Theta_{att})) \tag{3}$$

where $\sigma_2\ (x_i,\, c) = \frac{1}{1+\exp(x_i,c)}$ is the sigmoid activation function, attention factor $\alpha_i\ \epsilon\,[0,\, 1]$, and $\sigma_1\ \left(x_{i,c}^l\ \right) = \max(0, x_{i,c}^l\ )$, whose feature map $x$ is obtained at the output of layer $l$, $i$ and $c$ denote spatial and channel dimensions, respectively, and the gating vector for each pixel $i$ is given as $g_i \epsilon R^{F_g}$ to determine focus regions. The gating vector contains contextual information to prune lower-level feature responses. Attention gate has a parameter set $\Theta_{att}$, containing linear transformations $W_x\,\epsilon\,R^{F_l\ X\ F_{int}}$, $W_g\,\epsilon\,R^{F_g\ \ X\ F_{int}}$, $\Psi\,\epsilon\,R^{\,F_{int}\ X\ 1}$ and bias terms $b_\Psi\,\epsilon\,R$, $b_g\,\epsilon\,R^{F_{int}}$, where $R^{F_g\ \ X\ F_{int}}$ provides contextual information to guide the attention mechanism regarding where to look for salient features, $F_g$ is the coarse gating signal from the decoder layer, the concatenated features $x^l$ and g are linearly mapped to a $R^{F_{int}}$ dimensional intermediate space.

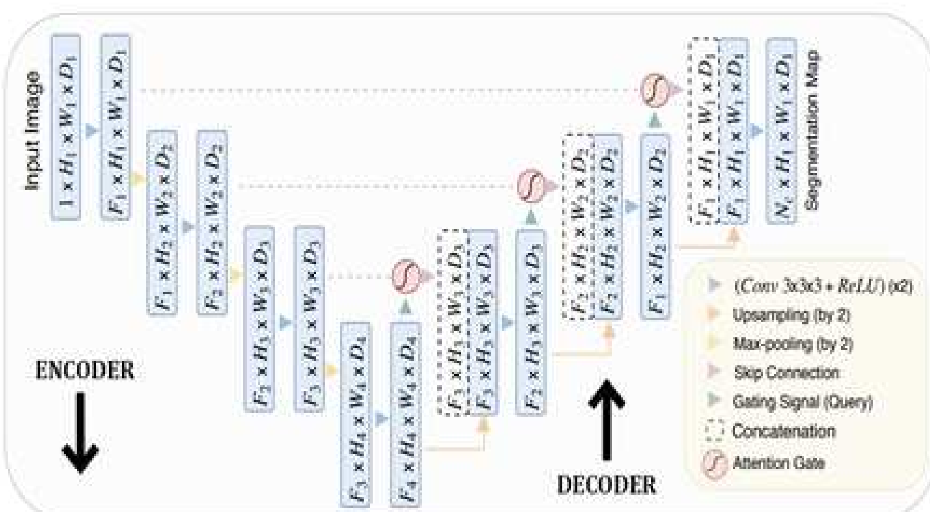


Figure 4. Attention U-Net model structure, Oktay et al. (2018)

Attention U-Net has been successfully applied to problems such as ocular disease diagnosis, lung cancer, abdominal structure segmentation, fetus development, brain tissue quantification and thyroid nodule classification, see Siddique et al. (2021) and Wang et al. (2023).

### 3.5. Recurrent U-Net

Recurrent U-Net is an extended version of U-Net, which adds a recurrent unit over some of its encoder and decoder paths. This enables the network to keep track of and to iteratively update more than just a single-layer internal state and provides for the flexibility to choose the portion of the internal state that we exploit for recursion purposes, Wang et al. (2019). Recurrent U-Net can operate efficiently in environments with limited training data and computational power. The recurrent U-Net model is depicted in Fig. 5.

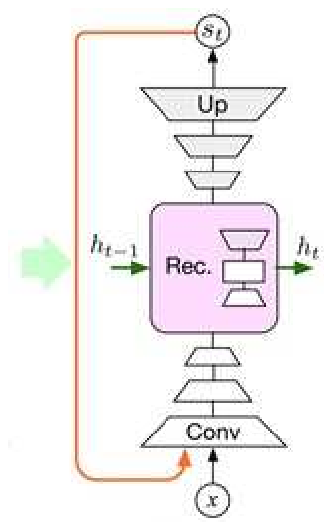


Figure 5. Recurrent U-Net, Wang et al. (2019)

The figure illustrates the architecture of a Recurrent U-Net (R2U-Net) block, which integrates convolutional and recurrent processing for improved feature learning in medical image segmentation. The input $x$ is first processed by a convolutional block to extract spatial features. These features, along with the hidden state from the previous time step, $h_{t-1}$, are passed into a recurrent unit (*Rec.*), which refines the features by incorporating temporal or iterative context. The updated hidden state $h_t$ is then moved forward, supporting sequential learning. Following this, an up sampling block (*Up*) increases the spatial resolution to reconstruct the output segmentation map $s_t$. The orange arrows represent skip connections that preserve spatial details from earlier layers, while the green arrows depict the recurrent state transitions, enabling the model to learn from previous iterations. This design allows the network to iteratively refine its predictions and is especially effective for handling complex or noisy medical images.

### 3.6. Residual U-Net

Residual U-Net is a variation of the U-Net architecture that incorporates residual connections between corresponding layers in the encoder and decoder paths, He et al. (2016). These connections allow the network to bypass some layers and propagate the information from the encoder directly to the decoder, so as to alleviate the vanishing gradients problem. The main advantage of using residual connections in U-Net is that they help to train deeper networks by preventing performance loss when adding more layers. The skip connections allow the network to access lower-level feature maps at different depths of the architecture, which can improve the overall segmentation performance. The residual U-Net can be implemented by modifying the basic U-Net architecture through adding skip connections between corresponding layers in the encoder and decoder paths. These connections can be implemented by element-wise addition or concatenation, depending on the design choice. The architecture of a residual U-Net can vary depending on the specific application and dataset. The residual learning variant is depicted in Fig. 6. In this residual learning framework, the stacked layers are not directly fit into the network, instead, these layers fit a residual mapping. If the desired mapping is $H(x)$, let the stacked nonlinear layers fit another mapping of

$$F(x) = H(x) - x\,. \tag{4}$$

The original mapping is reorganized into $F(x)+x$. It can be seen that optimizing the residual mapping is easier than optimizing the original mapping. If an identity mapping is optimal, the residual is simply pushed toward zero rather than forcing a stack of nonlinear layers to approximate an identity mapping. The function $F(x) + x$ can be realized by feed forward neural networks with shortcut connections, He et al. (2016). Several versions of Residual U-Net, like MGB-U-Net, have been developed for better segmentation and classification. They improve on traditional methods by using a Bottleneck Transformer (BOT) and a Multi-Conv Ghost Switchable Residual Block, see Hussein et al. (2024).

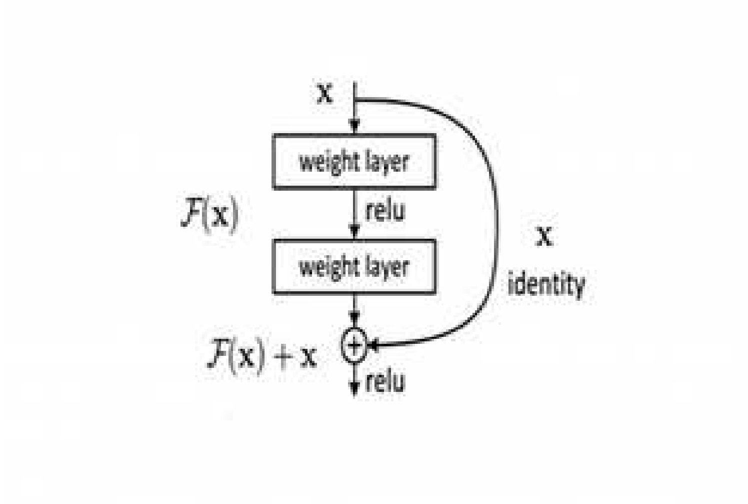


Figure 6. Residual learning variant, He et al. (2016)

### 3.7. Dense U-Net

In Dense U-Net, the idea of direct dense connections from the previous layers to the present layers is introduced, so that information flows efficiently between layers. Each layer obtains features from all previous layers, Wang et al. (2019). This model can eliminate the vanishing gradient problem and improve network performance, compared to Residual U-Net, whose performance decreases with the increase in the number of layers. The structure of the Dense U-Net model is demonstrated in Fig. 7.

The output of the $n^{th}$ layer in a dense block can be formulated as, Wang et al. (2019):

$$R_n = F_n([R_0,R_{1,}, R_{2,}, \dots R_{n-1}]), \tag{5}$$

where $[R_0,R_{1,}, R_{2,}, \dots R_{n-1}]$ denotes the concatenation of feature maps, produced in layers 0, 1,... $n$-1 and $F_n$ includes batch normalization, ReLU and convolution operations.

### 3.8. R2U++

R2U-Net improves feature extraction by making the network deeper, but it does not deal with the semantic gap problem. R2U++ overcomes these issues

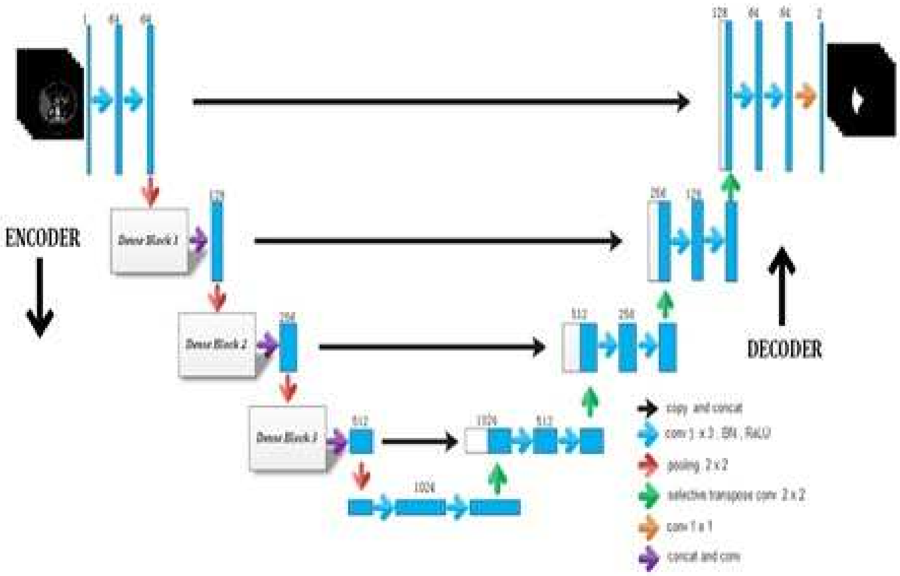


Figure 7. Dense U-Net, Wang et al. (2019)

in medical image segmentation architecture by using a deeper recurrent residual convolution block to extract crucial features for segmentation and dense skip pathways to accumulate features coming from multiple layers and apply concatenation accordingly, Mubashar et al. (2022).

The skip pathways can be computed as in equation (6) below. Namely, if $m$ is the index of down sampling of encoder layer, $n$ is the index of convolution layer in the skip pathways, the concatenated input to convolutional layer $X^{(m,n)}$ is given by equation (6), see Mubashar et al. (2022):

$$X_i^{m,n} = \left\{ \begin{array}{l} X^{m-1,n},\ n = 0 \\ \left[\left[X^{m,k}\right]_{k=0}^{n-1},\ u\left(x^{m+1,n-1}\right)\right],\ n > 0 \end{array} \right\} . \tag{6}$$

The feature map for the $X^{(m,n)}$ convolutional layer is

$$X_0^{m,n} = H\left(x_i^{m,n}\right) , \tag{7}$$

where $H(.)$ represents the recurrent residual convolution, $u(.)$ represents the upsampling from lower level, and the large square brackets indicate the concatenation function. Figure 8A shows the architecture of R2U++ with $L^4$ depth, and Fig. 8B shows the Ensemble R2U++.

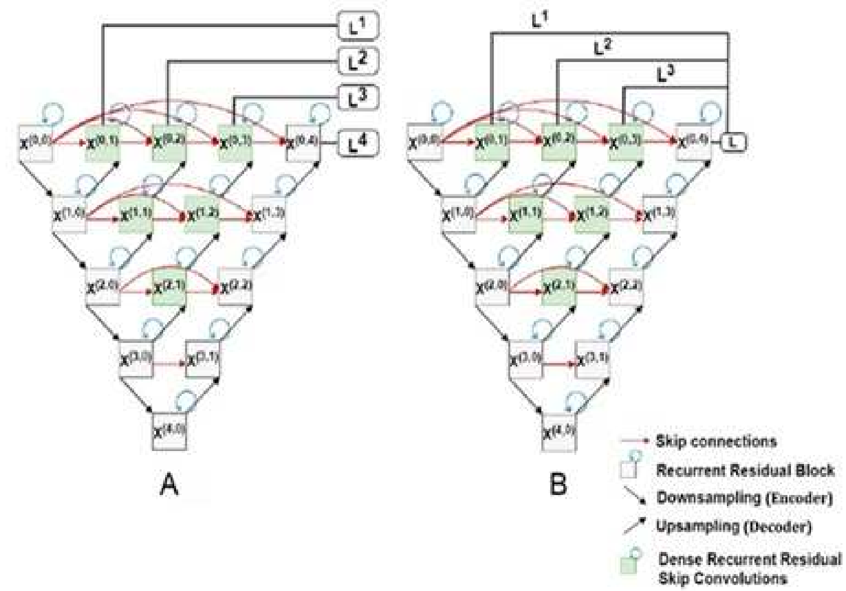


Figure 8. A. R2U++ with L4 depth B. Ensemble R2U++, see Mubashar et al. (2022)

### 3.9. MEDU-NET+

Most of the versions of U-Net, discussed before, improved the basic convolution block, added residual connections between convolution blocks, or deepened and widened the network. These methods, however, lead, as a rule, to an explosive increase in the number of parameters. MEDU-Net+ network integrates the information between encoder and decoder, and also improves the encoder, decoder and skip connection for getting better performance with a moderate increase in parameters, Yang et al. (2024). The architecture of MEDU-NET+ is depicted in Fig. 9.

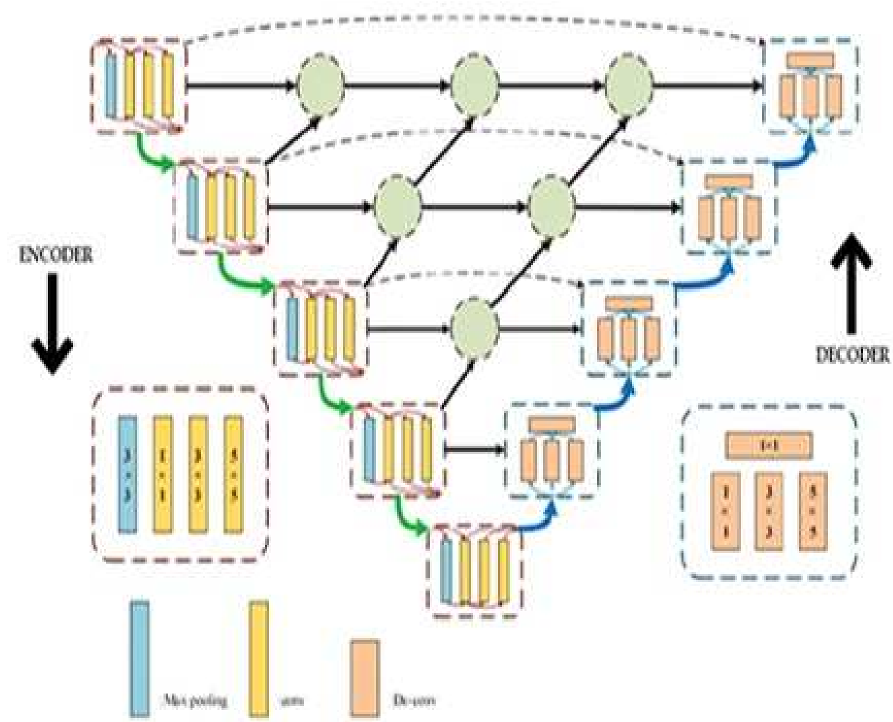


Figure 9. MEDU-NET+, Yang et al. (2024). MEDU-NET+ architecture contains the multi-scale encoding combined with GoogLeNet, the new layer-by-layer skip connection and the multi-scale feature fusion of the decoder. Green nodes in the figure denote attention-enhanced convolutional blocks used in the MEDU-Net+ architecture for improved feature learning)

The MEDU-Net+ network introduces some key improvements, intended to enhance medical image segmentation, including Enhanced Encoder with Inception Module, Layer-by-Layer Skip Connection and Multi-Scale Feature Fusion in the Decoder. The basic U-Net encoder is replaced with a GoogLeNet-inspired inception structure, consisting of four branches, namely 1x1 convolution, concatenated 1x1 and 3x3 convolutions, concatenated 1x1 and 5x5 convolutions and 3x3 max pooling, followed by 1x1 convolution. The outputs of these branches are concatenated, so as to create a rich multi-scale feature representation. A new skip connection mechanism is introduced, functioning as a cascade of multiple U-shaped networks. This connection transfers features through convolution, deconvolution, and aggregation, ensuring better semantic information retention. The decoder integrates multiple receptive fields by applying 1x1, 3x3, and 5x5 deconvolutional filters. These are aggregated to improve segmentation accuracy, ensuring more comprehensive feature recovery. Overall, MEDU-Net+ maintains a single structured architecture, eliminating redundant branches, making it flexible and modular for future enhancements.

### 3.10. U-Net with Explainable AI

The key focus of U-Net with Explainable AI is on enhancing explainability using Local Interpretable Model-Agnostic Explanations (LIME), Lakshmi et al. (2025). The Local Interpretable Model-Agnostic Explanations (LIME) technique generates visual interpretations of model predictions, ensuring transparency in clinical applications. The description of LIME is given by equation (8), see Lakshmi et al (2025):

$$\xi(x) = \arg\min_{g \in G} L(f, g, \pi_x) + \Omega(g) \tag{8}$$

where $L(f, g, \pi_x)$ measures fidelity, $\Omega(g)$ is a regularization term that controls complexity, $f$ denotes the original complex (black-box) model, $g \in G$ is an interpretable surrogate model that approximates $f$ in the local vicinity of the input instance $x$, and $\pi_x$ is a proximity kernel that assigns higher weights to samples closer to $x$, ensuring that the explanation focuses on the local behavior of $f$. Interpretation can be defined as a method with various justification fidelity functions, families, and complex measures. LIME highlights the most influential regions in MRI images, making model-generated decisions interpretable to radiologists.

### 3.11. Other hybrid U-Net models

Besides the aforementioned U-Net variants, there are other hybrid U-Net models, these extensions including ASU-Net++, see Gao and Almekkawy (2021), DSU-Net, see Wang et al. (2023), NU-Net, see Chen et al. (2023a), MMSA-Net, see Li et al. (2022), FE-HU-NET, see Nizamani et al. (2023), and U-Net variant integrating the Inception–Res block and the Dense Inception block, see Zhang et al. (2020), each incorporating specialized architectural enhancements to address specific challenges in medical image segmentation.

And so, ASU-Net++ is a modified version of U-Net++ architecture, enhanced with dilated short skip connections and adaptive ASPP layers to improve segmentation accuracy, especially for complex tumor edges. It leverages deep supervision, multi-scale context, and the AdaBelief optimizer to achieve high performance and fast convergence on ultrasound and CT liver tumor datasets, Gao and Almekkawy (2021). DSU-Net (Deep Supervision U-Net), a variant of the U-Net model, is designed to enhance segmentation in US images with the improved feature of deformable convolutions, Wang et al. (2023).

NU-Net is a powerful nested U-Net for accurate segmentation of tumors, Chen et al. (2023a). NU-net uses U-Nets with different depths and shared weights to better capture tumors at varying scales, and adds multi-step short-connections to strengthen long-range feature correlation – all without excessive architectural complexity. MMSA-Net (Multi-modal and Multi-scale Adversarial Segmentation Network) is another variant of U-Net, which was designed for segmentation from unregistered multi-parametric MRI data. The architecture consists of shared encoder, dual decoder and adversarial learning modules, Li et al. (2022).

FE-HU-Net is a hybrid U-Net variant, designed for medical image segmentation, regarding ultrasound images, Nizamani et al. (2023). This model incorporates a dual-encoder design, combining traditional convolutional layers with a feature-enhanced encoder, based on a residual attention mechanism. Yet another U-Net variant incorporates Inception-Residual blocks in the encoder path for capturing multi-scale features efficiently, and Dense Inception blocks in the decoder path to improve feature reuse and gradient flow. Together, these modules boost segmentation accuracy, especially in complex medical imaging tasks, outperforming standard U-Net in experiments, carried out, Zhang et al. (2020).

Table 1 provides a summary opinion on the here listed variants of the U-Net methodology.

## 4. Application of U-Net architectures for multi-organ segmentation in ultrasound and MRI imaging modalities

### 4.1. US/MRI liver images

U-Net and its variants have been successfully applied for the accurate segmentation and analysis of ultrasound and MRI liver images.

Table 1. Summary of U-Net variants and their key features

| U-Net variant | Key features |
|---|---|
| Basic U-Net | Encoder-decoder architecture with skip connections; basic for semantic segmentation tasks. |
| 3D U-Net | 3D Convolutions, 3D Feature maps and Encoder-Decoder structure adapted for 3D image segmentation |
| U-Net++ | Nested skip pathways, dense connections reducing the semantic gap between the feature maps of the encoder and decoder subnetworks. |
| Attention U-Net | Integrates attention gates to focus on relevant spatial features, enhancing segmentation of small or complex regions. |
| Recurrent U-Net | Incorporates recurrent convolutional layers (RNN-like) for temporal or contextual dependency in features. |
| Residual U-Net | Adds residual connections within encoder/decoder blocks to ease training and allow deeper networks. |
| Dense U-Net | Uses dense connectivity (from DenseNet), where each layer receives inputs from all previous layers in the block, improving feature reuse |
| R2U++ | Combines recurrent and residual units in a U-Net++ framework to capture spatial and temporal context with deep supervision. |
| MEDU-NET+ | Replaces encoding blocks with Inception modules; enhances multi-scale feature extraction while maintaining U-Net's overall structure. |
| U-Net with Explainable AI | Integrates explainability techniques (e.g., Grad-CAM, SHAP) to visualize and interpret segmentation decisions. |
| Other Hybrid U-Net models | ASU-Net++: Uses adaptive short skip connections with ASPP modules; emphasizes multi-scale context and deep supervision.<br>DSU-Net: Applies deep supervision across decoder stages; improves training efficiency and segmentation accuracy.<br>NU-Net: Nested U-Nets with shared weights; captures multi-scale features while keeping model efficient.<br>MMSA-Net: Designed for unregistered multi-modal MRI; features shared encoder, dual decoders, and adversarial multi-scale supervision.<br>FE-HU-Net: Combines dual encoders (standard + attention-based) with hybrid skip connections and channel attention for ultrasound image segmentation.<br>DIU-Net (Inception + Dense): Employs Inception-Res blocks in encoder and Dense Inception blocks in decoder for multi-scale and deep feature representation. |

A modified version of U-Net++, named ASU-Net++, uses adaptive nested U-Net and the short skip connections is used for the segmentation of ultrasound liver images, Gao and Almekkawy (2021). Atrous Spatial Pyramid Pooling (ASPP) is modified to an adaptive pooling structure for better compatibility with nested U-Net and for extracting features from different levels. For faster convergence rate, AdaBelief optimizer is used. Experiments were conducted to evaluate improved performance of this model in accurately segmenting tumors of varying sizes with complex edges. The results from experiments showed that the proposed model outperformed multiple network structures, and achieved a Dice coefficient of 0.915 for the ultrasound dataset. Compared with the original U-Net, using Adam optimizer, this model has a Dice coefficient increase of 8.4%, IoU increase of 10.8%, and MSE decrease of 45.9%. Compared with the original Attention U-Net, this model has a Dice coefficient increase of 5.3%, IoU increase of 9.1%, and MSE decrease of 36.5%. The challenge for this segmentation model is the speckle noise present in ultrasound images, making the images more difficult to segment.

An automatic liver parenchyma segmentation for MRI images was proposed using modified U-Net architecture, see Prasad et al. (2021), which included reduced convolutional layers. Dropout layers were used to address the problem of overfitting. Experiments were performed on the 3D-IRCADb dataset using images from 20 subjects, and the Dice score achieved was 0.945, indicating performance that was improved over the compared approaches.

### 4.2. US/MRI prostate images

An enhanced version of the U-Net model, called DSU-Net, has been proposed for prostate segmentation in ultrasound images, Wang et al. (2023). The improved quality in this model is the integration of shear transformation and deformable convolution, which makes the model more sensitive to border features and hence producing more effective segmentation. Experiments performed with 1057 prostate TRUS images from Shanghai Ruijin Hospital showed that DSU-Net has higher Dice coefficient (0.957) than other recent segmentation methods. The proposed model solved the problem of inaccurate segmentation results, caused by irregular prostate shape. However, the unclear boundary regions still need further investigation.

An efficient prostate image segmentation for MRI images was implemented using U-Net based model, with residual modules, Chahal et al. (2021). Preprocessing methods were employed to improve the performance. Experimentations were performed on Promise12 MRI dataset of 50 subjects and compared to results from various U-Net based models to evaluate the performance. With simple curvature preprocessing, the proposed method obtained the Dice coefficient value of 0.9453. With CLAHE-0.01(ClipLimit), Dice score was 0.9531, with CLAHE-0.05(ClipLimit) preprocessing, Dice score value was 0.9481, and for histogram equalization preprocessing, Dice score was 0.9453. The proposed model outperformed the other selected models, achieving an average Dice score of 97.50%, which was higher than VGG19RSeg with average Dice score of 94.57%, Active Appearance Model (AAM) with average Dice score of 87%, Long FCN with 85.3% average Dice score, CNN with 78% average Dice score, and VGG19-equipped FCN with 87.7% average dice score. The results of experiments carried out support the conclusion that the proposed model performs better than the other methods taken for study.

### 4.3. US/MRI pancreas images

Attention U-Net was successfully employed for the automatic segmentation of the pancreatic cyst lesions, Oh et al. (2021), on endoscopic ultrasonography (EUS) images. The performance of Attention U-Net was compared with the basic U-Net, Residual U-Net, and U-Net++ models. Attention U-Net produced a better Dice similarity coefficient and intersection over union scores than the other models on the internal test. But no significant difference was observed between Attention U-Net and Residual U-Net or between the Attention U-Net and U-Net++. Attention U-Net showed the highest DSC and IoU scores, equal to 0.794 and 0.741, respectively. U-Net++ showed the lowest performance indicators, with DSC of 0.727 and IoU of 0.628. Attention U-Net had the highest pixel accuracy of 0.983 and yielded the highest recall score at IoU $>$ 0.70 and IoU $>$ 0.85 (0.857 and 0.571, respectively).

Segmentation of MRI pancreatic images was performed with multimodal and multi-scale adversarial network (MMSA-Net), Li et al. (2022). MMSA-Net consists of two complementary modules of a shared encoder and dual decoder. The inter-modality adversarial learning helps to combine shared features from different modalities by making the features extracted by the encoder more similar. The intra-modality multi-scale adversarial supervision focuses on preserving each modality's unique features while still using the shared ones, ensuring accurate segmentation results for each modality. Experiments conducted on MRI pancreatic images show that the proposed method improves the performance of segmentation. This is evident from the improved Dice value as compared to the existing pancreatic segmentation methods, which obtained an average Dice value of less than 64%.

### 4.4. US/MRI brain images

An automatic brain ultrasound image segmentation method was proposed for delineating tumor resection cavities in intraoperative ultrasound (iUS) images to aid image registration during neurosurgery, utilizing both 2D and 3D U-Net network architectures for training and evaluation, Carton et al. (2020). Experiments were conducted on 2-D RESECT and 3-D BITE ultrasound image datasets of 37 and 13 cases, respectively. 3D U-Net performed better, with 0.72 mean and 0.88 median Dice score, than 2D U-Net with a median Dice score over 0.8. This improved performance motivates further studies with more clinical datasets, for automated analysis of the US tumor images. Here the proposed model gained improved performance with a limited dataset and low quality of US images, but a better validation would have been possible with a larger dataset.

A hybrid U-Net architecture integrated with Inception-Res block and Dense-Inception module is proposed for efficient MRI brain image segmentation, Zhang et al. (2020). This model extracts features without additional parameters. Down-sampling block reduces the size of feature maps to accelerate learning and the up-sampling block resizes the feature maps. Experiments were performed on images of blood vessel segmentations from retina images, the lung segmentation of CT data from the benchmark Kaggle datasets and the MRI scan brain tumor segmentation datasets from MICCAI BraTS 2017. The results from experiments showed that the proposed method provides better performance compared to the other models considered. For lung segmentation, average Dice score is 0.9857, for blood vessel segmentation, average Dice score is 0.9582, and for brain tumor segmentation, average Dice score is 0.9867. The limitation of the proposed DIU-Net model is that in the Dense-Inception block increasing the growth rate may lead to too many parameters, making the model more difficult and slower to train.

Automatic Brain MRI segmentation is proposed using Attention residual U-Net, Zhang et al. (2020), architecture, which integrates attention units and residual units into U-Net. To address the class imbalance problem, a combined loss function of weight cross-entropy and generalized Dice loss is adopted. Experiments were conducted on two MRI brain tumor segmentation datasets, BraTS 2017 and BraTS 2018. The results showed improved performance of AResU-Net over benchmark methods, including base U-Net, FCNN, ResU-Net, Densely CNN and the CNN. AResU-Net achieved mean Dice scores of 0.892, 0.853 and 0.825 on the whole tumor, core tumor and enhancing tumor, respectively. AResU-Net attains 6.1%, 5.2% and 7.5% gains over base U-Net. It gains 1.2%, 0.3% and 0.5% over ResU-Net. Moreover, compared with FCNN, CNN and Densely CNN, AResU-Net achieves the highest Dice score on the whole tumor and enhancing tumor segmentation. The advantage with the proposed method is that it shows better performance even with a limited dataset.

A hybrid U-Net variant named FE-HU-NET was proposed for enhancing the precision of brain MRI segmentation, Nizamani et al. (2023). Initially, they used image enhancement techniques CLAHE, MHE, and MBOBHE for each model. Next, they modified U-Net model with a personalized layer design. Finally, a CNN model is used for post-processing to refine the segmentation outcomes. Thus, HU-Net layer combined a tailored UNet layer and a CNN. Experiments were conducted on two public datasets, BraTS 2020 dataset and LGG-MRI segmentation dataset to demonstrate that accuracy rates of more than 99% are attained. Performance metrics including Jaccard index, sensitivity, and specificity, further validate the outstanding performance of this hybrid UNet models. However, there are challenges, like limited generalizability across different datasets, dependency on data quality and annotation, high demand of computational resources and ethical consideration.

### 4.5. US/MRI breast images

A simple nested U-Net (NU-Net) for accurate and efficient segmentation utilizes U-Nets with different depths and shared weights to achieve robust classification, Chen et al. (2023a). In this model, the 15 layer deep U-Net is taken as the backbone network to extract more telling breast tumor features. A multi-output U-Net is taken as the bond between the encoder and the decoder to enhance the network adaptability for tumors with different scales. Lastly, the short-connection based on multi-step down-sampling improves the correlation of long-range information of encoded features. Experiments performed on three public breast ultrasound datasets show that this method has a more competitive segmentation performance on breast tumors than baseline U-Net and other U-Net variants, including SegNet, Attention U-Net, U-Net++ , U-Net3+, SKNet and RDAU-Net.

The Ensemble Deep Convolutional Neural Network (EDCNN) model, featuring Explainable Artificial Intelligence (XAI) through the integration of Grad-CAM, demonstrated exceptional improvement in breast cancer detection from ultrasound images by enhancing both interpretability and diagnostic performance, see Islam et al. (2024). The authors thereof integrated MobileNet and Xception models and employed standard preprocessing, data normalization and data augmentation techniques. Experimentation and evaluation of the proposed model demonstrated accuracy of 87.82% on BUSI dataset and 85.69% on UDAIT dataset.

An automated method for breast cancer detection, based on modified U-Net architecture for Dynamic Contrast Enhanced Magnetic Resonance Images was proposed by Khaled et al. (2022). This model is an ensemble method combining three U-Net models, each using a different input combination, outperforming all individual methods and other existing approaches. For experimentation, a subset of 46 cases from the TCGA-BRCA dataset was used. Due to the incomplete annotations provided, they were complemented with the help of a radiologist in order to include secondary lesions that were not originally segmented. The proposed method attains the mean Dice similarity coefficient of 0.680 (0.802 for main lesions), which is better than for the state-of-the art methods using the same dataset, demonstrating the effectiveness of that method considering the complexity of the dataset.

At this point we close the proper review on various U-Net models used for the segmentation of ultrasound and MRI images.

## 5. Performance metrics

The performance of various U-Net medical image segmentation models is being validated quantitatively using various evaluation metrics, including Dice coefficient, accuracy, specificity, precision, recall, Jaccard index, and F-score (see Siddique et al., 2021; Chen et al., 2023a; Kumar et al., 2018b). Let TP be the total number of lesion pixels that are common in both segmented image (SI) and ground truth (GT), TN be the total number of the pixels of the background that are common in both SI and GT, FN the total number of pixels that are actually in the lesion, but not shown in the segmented region, and FP the total number of pixels, which are not actually in the lesion, but shown in the segmented region. These basic numbers shall be used in the definitions of the here considered performance metrics.

*Dice* coefficient (DC), compares the machine segmented and ground truth binary image and its value ranges from 0 (no matching) to 1 (perfect matching), Kumar et al. (2018b). The *accuracy* of segmented image (SI) with respect to ground truth image (GT) gives a measure of the overall efficiency of the algorithm, Siddique et al. (2021), and reaches the absolutely best value at 1 or 100. The *specificity* gives the percentage of the correctly extracted region of interest by the segmentation algorithm, Siddique et al. (2021), and reaches its best at 1 or 100%. *Precision* is a function of true positives and samples misclassified as positives (false positives). It gives a measure of the predictive power of the algorithm and reaches also its best at 1 or 100%. The precision of the extracted region of interest by the proposed method can be calculated as in Table 2. The *recall* is a function of its correctly segmented samples (true positives) and its misclassified samples (false negatives). The recall takes values in the range [0, 1] or [0, 100] and reaches its best at 1 or 100%, Kumar et al. (2018b).

The Dice coefficient and Jaccard index are proportional to the amount of spatial overlap between machine segmented and ground truth binary image and so their values range from 0 (no overlap) to 1 (perfect matching), Kumar et al. (2018b). The relationship between DC and JC is provided in Table 2. *F-score* is used to measure the overall performance of a segmentation model by combining precision and recall, Siddique et al. (2021). The mathematical expressions for the performance metrics, used in evaluation of segmentation models, are given in Table 2.

Table 2. Performance metrics for the evaluation of segmentation models

| ***Performance Metrics*** (*TP-True Positive, FP-False Positive, TN-True Negative and FN-False Negative Pixels*) |
|---|
| $Dice\,Coefficient\,(DC) = \frac{2\,\lvert TP\rvert}{2\,\lvert TP\rvert+\lvert FP\rvert+\lvert FN\rvert}$ |
| $Accuracy = \frac{TP+TN}{TN+FN+TP+FP}$ |
| $Specificity = \frac{\lvert TN\rvert}{\lvert TN\rvert+\lvert FP\rvert}$ |
| $Precision = \frac{\lvert TP\rvert}{\lvert TP\rvert+\lvert FP\rvert}$ |
| $Recall = \ Sensitivity = \frac{\lvert TP\rvert}{\lvert TP\rvert+\lvert FN\rvert}$ |
| $IoU = Jaccard\,Index = \frac{DC}{2-DC}$ |
| $F-score = \frac{2*Precision*Recall}{Precision+Recall}$ |

## 6. Results and discussion

This review article outlined and then discussed several variants and improvements of the basic U-Net architecture, proposed by various researchers and practitioners. Table 3, which follows now, summarizes the variants of U-Net models used in medical image segmentation, including exemplary performance indicators.

Table 3 shows the comparison of performance of different U-Net variants used in medical image segmentation. From the table, it can be seen that the performance of the particular variants of U-Net model differs according to the applications domain and segmentation requirements. New variants of U-Net models were introduced to address challenges of the existing U-Net models.

Table 3: Comparison of performance of various U-Net models used in medical image segmentation

| **U-Net model** | **Dataset** | **Modality** | **Performance evaluation parameter** |
|---|---|---|---|
| Basic Net, Ronneberger, Fischer & Brox (2015) | ISBI Challenge (PhC-U373), ISBI(DIC-HeLa) | Microscopic images | Dice Coefficient 0.9203 0.7756 |
| 3D U-Net, Zhou et al. (2020) | BRATS 2020 Brain | MRI | Accuracy 0.994 IoU 0.8023 |
| U-Net++, Zhou et al. (2018) | Data Science Bowl 2018, ASU-Mayo [PubMed: 26978662] MICCAI 2018 LiTS, LIDC-IDRI [PubMed: 21452728] | Microscopy, RGB video, CT, CT | IoU 0.9263, 0.3345, 0.8290, 0.7721 |
| Attention U-Net, Oktay et al. (2018) | Real Pancreas CT-150 images, NIH-TCIA | CT | Dice coefficient 0.831 |
| Recurrent U-Net DRU, SRU, Wang et al. (2019) | Hand dataset (Egohands, EYTH, HOF, KBH and GTEA) Retina DRIVE | Camera, Fundus | Dice coefficient 0.898 0.902 0.889 0.957 0.959 0.821 |
| Residual U-Net, He et al. (2016) | ImageNet 2012 | Camera | mAP 83.8 |

| | | | |
|---|---|---|---|
| Dense U-Net, Wang et al. (2019) | 3D-IRCADb, LITS | CT | Dice coefficient<br>0.949<br>0.921 |
| R2U++, Mubashar et al. (2022) | EM ISBI 2012-, COVID-19 CT, DRIVE JSRT | Electron microscopy, X-rays, fundus, and CT | Dice coefficient<br>0.949<br>0.772<br>0.941<br>IoU<br>0.904<br>0.628<br>0.941 |
| MEDU-NET+, Yang et al. (2024) | DRIVE, ISBI2012, CHAOS | Retina, Electron microscopy, MRI | Accuracy<br>0.959<br>0.783<br>0.790<br>IoU<br>0.954<br>0.786<br>0.789<br>MIoU<br>0.955<br>0.782<br>0.780 |
| U-Net with Explainable AI, Lakshmi et al. (2025) | BT-MRI | MRI-Brain | Accuracy 0.978<br>Precision 0.954<br>Recall 0.954<br>F1-score 0.955<br>MCC 0.939 |

The Basic U-Net model architecture, Ronneberger, Fischer and Brox (2015), has the merit of faster segmentation with GPU (Graphical Processing Unit) systems, but with a limited dataset. The small size of the dataset leads to overfitting, where the model memorizes the training samples rather than learning generalized patterns. A method to handle overfitting is the application of dropout during the training process, where some neurons are eliminated in each iteration. Also, U-Net does not, of course, use innovative mechanisms, like attention or recurrent units, for improvement of performance in some applications.

3D U-Net extends the U-Net architecture by introducing volumetric segmentation and replacing all 2D operations with their 3D counterparts, Sangui et al. (2023). This helped to analyze the 3D MRI images more efficiently and is generally more suitable for medical image analysis. But the challenge of this approach is that medical images acquired from various medical equipment follow different standards and hence annotations or performance of CT/MRI machines are not uniform. Also, computational cost is higher due to the requirement of more memory and long training times.

An improved version of U-Net was U-Net++, presented by Zhou et al. (2020), where nested and dense skip connections were added for medical image segmentation to reduce the semantic gap between the feature maps of the encoder and decoder sub-networks. The dense skip connections helped the model to capture fine-grained and complex features such as small lesions, subtle texture variations, and boundary details across multiple scales. However, this comes at the cost of increased computational requirements and a higher risk of overfitting if not properly regularized.

A modified version of U-Net, named Attention U-Net, introduced attention mechanisms to focus on important features in the relevant regions, while ignoring unnecessary areas so as to improve segmentation performance in noisy images, Oktay et al. (2018). This model improved the performance of network over the basic U-Net and U-Net++, especially for noisy background images, but presents additional computational overhead, making the model slower during training and inference.

Recurrent U-Net is an extended version of U-Net, which can operate efficiently in environments with limited training data and computational power, by adding a recurrent unit over some of its encoder and decoder paths, Wang et al. (2019). The performance of this model is better than that of the standard U-Net, especially for complex segmentation tasks that need contextual understanding over multiple layers or steps. Deep Recurrent Unit focuses on challenging datasets, while Spatial Recurring Unit focuses on spatial relations for better feature extraction.

Residual U-Net is a variation of the U-Net architecture that incorporates residual connections between corresponding layers in the encoder and decoder paths, He et al. (2016). It helps to eliminate the vanishing gradient problem in deeper networks.

Dense U-Net model, Wang et al. (2019), eliminates the vanishing gradient problem and shows improved network performance compared to Residual U-Net. This model showed improvements on complex datasets in terms of training stability and model convergence. It is also robust to overfitting and noise.

In Dense U-Net, the idea of direct dense connections from previous layers to present layers is introduced so that information flows efficiently between layers, allowing for efficient feature reuse, leading to better performance. This architecture is beneficial in complex tasks requiring the model to capture a diverse set of features. The benefits in segmentation accuracy, especially in difficult datasets, outweigh its additional computational burden. Also, this model eliminates the vanishing gradient problem and improves network performance compared to Residual U-Net, whose performance decreases with increase in the number of layers.

R2U-Net addressed the problem of simple feature extracting blocks by making the network deeper, but it does not deal with the semantic gap problem. R2U++ overcomes these issues in medical image segmentation architecture by using a deeper recurrent residual convolution block to extract crucial features for segmentation and dense skip pathways to accumulate features coming from multiple layers and apply concatenation accordingly, Mubashar et al. (2022), i.e., R2U++ combines the benefits of Residual U-Net, Recurrent U-Net, and UNet++. Even though the model involves high memory and computational cost, it can provide significant results in the case of complex applications.

The MEDU-Net+ network introduces some key improvements to enhance medical image segmentation of retina, microscopy and brain images using the modules such as Enhanced Encoder with Inception Module, Layer-by-Layer Skip Connection and Multi-Scale Feature Fusion in the Decoder, Yang et al. (2024). Further, U-Net with Explainable AI presents balanced performance in MRI-brain analysis, with accuracy of 0.978 and F1-score of 0.955, proving its reliability in medical imaging.

Attention U-Net, R2U++, and Dense U-Net provide superior performance for complex or noisy datasets at the cost of higher computational burden. 3D U-Net, R2U++, and Dense U-Net are more complex models and hence also computationally more expensive. Residual U-Net, Dense U-Net, and R2U++ models can be generalized better in more complex tasks.

In short, the strengths of particular U-Net architectures are mostly application dependent. 3D U-Net outperforms 2D U-Net when working with volumetric data such as MRI or CT scans, since it captures inter-slice contextual information, Attention U-Net works well with noisy data, and R2U++ is suitable for challenging segmentation tasks. Major challenge in segmentation is that the boundary between organs and abnormalities is difficult to be distinguished on the image and medical image segmentation requires exceptionally high accuracy for disease prediction. With the development and improvement of different models based on U-Net, including attention mechanism, dense module, residual module, U-Net have been used recently to achieve precise segmentation of lesions in various applications. However, still some challenges exist in the field of U-Net deep learning.

Table 4: Summary of application of U-Net models in US/MRI images (US: ultrasound)

| Dataset type | Model | Dice coefficient | Challenges |
|---|---|---|---|
| US Liver images, Gao & Almekkawy (2021) | ASU-Net++ | 0.915 | Speckle noise present in ultrasound images makes the images more difficult to segment |
| MRI Liver images, Prasad et al. (2021) | Modified U-Net | 0.945 | Experimented on a small 3D-IRCADb dataset |
| US Prostate images, Wang et al. (2019) | DSU-Net | 0.975 | Unclear boundary regions need further investigation |
| MRI Prostate images, Chahal et al. (2021) | Residual U-Net | 0.975 | Limited to a specific application |
| US Pancreas images, Oh et al. (2021) | Attention U-Net | 0.794 | Pixel properties of US images can be affected by hardware factors and speckle noise |
| MRI Pancreas images, Li et al. (2022) | MMSA-Net | 0.809 | Unregistered multi-modal data have varying morphologies, which may cause mutual interference between modalities during feature fusion, ultimately degrading segmentation performance |
| US Brain images, Carton et al. (2020) | 3D Net | 0.880 | The limited dataset and low-quality ultrasound images resulted in less reliable and less generalizable validation outcomes |

| | | | |
|---|---|---|---|
| MRI Brain images, Zhang et al. (2020a,b), Mubashar et al. (2022) | DIU-Net<br>Attention residual U-Net<br>FE-HU-NET(Hybrid U-Net) | 0.9867<br>0.892<br>0.99 | In the DIU-Net of Dense-Inception block increasing the growth rate may lead to too many parameters, making the model more difficult and slower to train.<br>Limited dataset<br>Limited generalizability across different datasets, dependency on data quality and annotation, high demand of computational resources and ethical considerations |
| US Breast Images, Chen et al. (2023a), Islam et al. (2024) | NU-Net<br>EDCNN (U-Net with XAI) | 0.90<br>0.878 | More adequate objective features can be extracted from complex ultrasound images using a deeper network framework.<br>Gray scale US images are used instead of RGB images. |
| MRI Breast Images, Khaled et al. (2022) | Ensemble U-Net | 0.802 | Incomplete annotations provided, complemented with the help of a radiologist in order to include secondary lesions that were not originally segmented. |

Table 4 summarizes the application of U-Net variants in ultrasound and MRI images. From Table 4, it is evident that FE-HU-NET (Hybrid U-Net) attained the best Dice score (0.99), excelling in MRI brain tumor segmentation, followed by Dense Inception U-Net across multiple datasets (0.9867). In FE-HU-NET, the precision is improved by using a tailored CNN and U-Net, whereas in DIU-Net, the vanishing gradient issue is reduced and feature extraction is improved using a combination of inception modules and dense connections. Recent advances in brain MRI includes U-net with Explainable AI, which showed a Dice score of 0.955, but with increased computational complexity.

ASU-Net++, with its nested U-Net structure and adaptive ASPP pooling, handled complex tumor edges efficiently in ultrasound liver tumor segmentation, accomplishing a Dice score of 0.915. U-Net Ensemble for breast lesions showed the effectiveness of ensemble learning, integrating three U-Net models with different input representations. The model achieved a competitive Dice score of 0.802 for main lesions, outperforming standalone U-Net variants in spite of the complex nature of breast MRI lesions. Brain Tumor Resection U-Net showed that 3D U-Net performed better than 2D U-Net in intraoperative ultrasound images, with a Dice score of 0.88. However, 2D U-Net had lower computational costs and faster inference times.

Dense inception U-Net and ASU-Net++ with better segmentation accuracy need high computational power due to their deeper architectures. U-Net Ensemble leverages multiple models, which increases training time and memory consumption, yet it offers better generalization across datasets.

DSU-Net for US prostate image segmentation improved boundary detection addressing challenges posed by poor prostate gland visibility in TRUS images. U-Net models for pancreatic cyst segmentation employed attention mechanisms to refine lesion segmentation, demonstrating the potential of deep learning in aiding early diagnosis.

The applications of the different variants of U-Net model, discussed above, demonstrate progress attained in the area of medical image segmentation. Still, there are some areas that could be improved and thereby enhance performance. Thus, the future scope of this research includes multimodal data integration, where the model can analyze images of different modalities, including US, CT and MRI. Models such as ASU-Net++ for liver ultrasound and Attention U-Net for pancreas ultrasound are limited by the inherent noise and pixel-level artifacts introduced during image acquisition by ultrasound hardware, which can degrade segmentation performance. Thus, the research can be extended to creating models, which are more generalized across different types of medical imaging, so as to improve their capability to handle noise and hardware factors. The next challenge is to enhance performance for small and low-quality datasets, like MRI brain images with DIU-Net. In this case, more sophisticated data augmentation and transfer learning techniques can be developed. Exploring advanced boundary detection techniques, like deep learning-based edge detection or integrating multi-scale information is necessary to handle the challenge of unclear boundary region, noted in DSU-Net model for US prostate images. More efficient hybrid models, like those combining U-Net with Attention mechanism, should be developed to tackle the problem of unclear boundary regions in segmentation. The DIU-Net, used in MRI brain images, highlighted the difficulty of training time, due to a large number of parameters. Future research could focus on optimizing network architectures, including reduction of the number of parameters or using lightweight networks to improve speed and efficiency in training. Pruning techniques or efficient architectures, like MobileNets or EfficientNet, adaptive or self-optimizing networks that adjust their complexity, depending on the input data, could be explored to improve model efficiency. For datasets like the MRI breast images and brain images, where incomplete annotations are a challenge, the integration of semi-supervised learning and active learning could be explored, where algorithms are used to ask for labels in uncertain areas, potentially reducing reliance on human annotations and improving segmentation accuracy. Developing more effective tools that allow radiologists to easily annotate images in collaboration with AI tools could help generate more comprehensive annotations, improving the quality of training data for future models.

As AI models are increasingly used in clinical settings, ensuring transparency, interpretability, and fairness in the decision-making process of these models is critical. Future work should therefore also involve developing methods for explainable AI to provide experts with insights into how models make decisions, ensuring trust and facilitating the adoption of AI technologies in medical practices. It can hence be summarized that while the models given in Table 2 showed impressive progress in medical image segmentation, addressing the identified challenges will lead to more robust, efficient, and clinically applicable models in the future. The selection of segmentation models has to be based on the nature of the application, complexity of dataset and the availability of the computational resources. Basic U-Net is a good option for applications with limited datasets, whereas advanced alternatives, like Dense U-Net, R2U++ and Attention U-Net give noteworthy results at the expense of high computational complexity.

## 7. Conclusion

This article reviews the application of different versions of U-Net in medical image segmentation. The variants discussed were Basic U-Net, 3D U-Net, U-Net++, Attention U-Net, Recurrent U-Net, Residual U-Net, Dense U-Net and R2U++. These models were meant to improve the accuracy and reliability of the previously developed approaches to automate the segmentation task. The choice of a specific variant depends on the application domain and segmentation requirements. Researchers should evaluate these variants to determine the best suited U-Net model. Advanced U-Net models continue to evolve, demonstrating remarkable improvements in medical image segmentation. Future research should focus on optimizing computational efficiency while maintaining high segmentation accuracy. The use of hybrid architectures, self-supervised learning, and transformer-based models could further enhance the capabilities of U-Net for medical applications.

## References

AHMAD, H. A., YU, H. J. AND MILLER, C. G. (2013) Medical Imaging Modalities. *Medical Imaging in Clinical Trials*, 3–26. https://doi.org/10.1007/978-1-84882-710-3_1

ALTAF, F., ISLAM, S. M. S., AKHTAR, N. AND JANJUA, N. K. (2019) Going Deep in Medical Image Analysis: Concepts, Methods, Challenges and Future Directions. *IEEE Access*, 7, 99540–99572

CARTON, F., CHABANAS, M., LANN, F. L. AND NOBLE, J. H. (2020) Automatic segmentation of brain tumor resections in intraoperative ultrasound images using U-Net. *Journal of Medical Imaging*, **7**(03), 1. https://doi.org/10.1117/1.jmi.7.3.031503

CHAHAL, E. S., PATEL, A., GUPTA, A., PURWAR, A. AND DHANALEKSHMI, G. (2021) U-Net based Xception Model for Prostate Cancer Segmentation from MRI Images. *Multimedia Tools and Applications*, **81**(26), 37333–37349. https://doi.org/10.1007/s11042-021-11334-9.

CHEN, G., LI, L., ZHANG, J. AND DAI, Y. (2023**a**) Rethinking the unpretentious U-Net for Medical Ultrasound Image Segmentation. *Pattern Recognition*, 142, 109728–109728. https://doi.org/10.1016/j.patcog.2023.109728.

CHEN, G., LI, L., DAI, Y., ZHANG, J. AND YAP, M. H. (2023**b**) AAU-Net: An Adaptive Attention U-Net for Breast Lesions Segmentation in Ultrasound Images. *IEEE Transactions on Medical Imaging*, **42**(5), 1289–1300. https://doi.org/10.1109/tmi.2022.3226268

CICEK, O., ABDULKADIR, A., LIENKAMP, S. S., BROX, T. AND RONNEBERGER, O. (2016) 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Lecture Notes in Computer Science*, **9901**, 424–432). Springer, Cham.

DU, G. ET AL. (2020) Medical Image Segmentation based on U-Net: A Review. *Journal of Imaging Science and Technology*, **64**(2). 020508–020512 Available at: https://doi.org/10.2352/j.imagingsci.technol.2020.64.2.020508.

EZUGWU, A. E., SHUKLA, A. K., AGBAJE, M. B., OYELADE, O. N., JOSÉ-GARCÍA, A. AND AGUSHAKA, J. O. (2020) Automatic clustering algorithms: a systematic review and bibliometric analysis of relevant literature. *Neural Computing and Applications*, **33**(11), 6247–6306. https://doi.org/10.1007/s00521-020-05395-4

GAO, Q.-H. AND ALMEKKAWY, M. (2021) ASU-Net++: A Nested U-Net with Adaptive Feature Extractions for Liver Tumor Segmentation. *Computers in Biology and Medicine*, **136**, 2021, 104688, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2021.104688.

He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep residual learning for image recognition. In: *Proc. IEEE Conf. Comput. Vis. Pattern Recognition (CVPR)*, Jun. 2016, 770-778.

Hesamian, M. H. et al. (2019) Deep learning techniques for Medical Image Segmentation: Achievements and Challenges. *Journal of Digital Imaging*, **32**(4), 582–596. Available at: https://doi.org/10.1007/s10278-019-00227-x.

Houssein, E. H., El-din Helmy, B., Oliva, D., Elngar, A.A. and Shaban, H. (2021) Multi-level Thresholding Image Segmentation Based on Nature-Inspired Optimization Algorithms: A Comprehensive Review. In: D. Oliva, E. H. Houssein and S. Hinojosa (eds.), *Metaheuristics in Machine Learning: Theory and Applications. Studies in Computational Intelligence*, **967**, 239–265, Springer, Cham. https://doi.org/10.1007/978-3-030-70542-8_11

Hussein, A., Youssef, S., Ahmed, M. A. and Ghatwary, N. (2025) MGB-Unet: An Improved Multiscale U-Net with Bottleneck Transformer for Myositis Segmentation from Ultrasound Images. *Journal of Imaging Informatics in Medicine*, 38, 217–228. https://doi.org/10.1007/s10278-024-01168-w

Islam, M. R., Rahman, M. M., Ali, M. S., Nafi, A. a. N., Alam, M. S., Godder, T. K., Miah, M. S. and Islam, M. K. (2024) Enhancing breast cancer segmentation and classification: An Ensemble Deep Convolutional Neural Network and U-Net Approach on Ultrasound Images. *Machine Learning with Applications*, 16, 100555. https://doi.org/10.1016/j.mlwa.2024.100555

Khaled, R., Vidal, J., Vilanova, J. C. and Martí, R. (2022) A U-Net Ensemble for Breast Lesion Segmentation in DCE MRI. *Computers in Biology and Medicine*, 140, 105093. https://doi.org/10.1016/j.compbiomed.2021.105093

Kumar, S. N., Lenin, F., Muthukumar, S, Kumar, A. and Varghese, S. (2018**a**) A Voyage on Medical image Segmentation Algorithms. *Biomedical Research*, Special Issue: S75-S87.

Kumar, S. N., Lenin, F. A., Ajay Kumar, H. and Varghese, P. S. (2018**b**) Performance Metric Evaluation of Segmentation Algorithms for Gold Standard Medical Images. In: P. K. Sa et al. (eds.), *Recent Findings in Intelligent Computing Techniques. Advances in Intelligent Systems and Computing*, **709.** Springer Nature Singapore, 457–469. https://doi.org/10.1007/978-981-10-8633-5_45

Lakshmi, K., Amaran, S., Subbulakshmi, G., Padmini, S., Joshi, G. P. and Cho, W. (2025) Explainable Artificial Intelligence with U-Net Based Segmentation and Bayesian Machine Learning for Classification of Brain

Tumors using MRI Images. *Scientific Reports*, **15**(1), 690. https://doi.org/10.1038/s41598-024-84692-7

Li, J., Feng, C., Shen, Q., Lin, X. and Qian, X. (2022) Pancreatic cancer segmentation in unregistered multi-parametric MRI with adversarial learning and multi-scale supervision. *Neurocomputing*, 467, 310–322 https://doi.org/10.1016/j.neucom.2021.09.058

Mubashar, M., Ali, H., Grönlund, C. et al. (2022) R2U++: a multiscale recurrent residual U-Net with dense skip connections for Medical Image Segmentation. *Neural Computing & Applications* 34, 17723–17739. https://doi.org/10.1007/s00521-022-07419-7

Nizamani, A. H., Chen, Z., Nizamani, A. A. and Bhatti, U. A. (2023) Advanced brain tumor segmentation using feature fusion methods with deep U-Net model with CNN for MRI data. *Journal of King Saud University - Computer and Information Sciences*, **35**(9), 101793-101806. https://doi.org/10.1016/j.jksuci.2023.101793

Oh, S. H., Kim, Y., Park, Y. and Kim, K. G. (2021) Automatic pancreatic cyst lesion segmentation on EUS images using a Deep-Learning approach. *Sensors*, **22**(1), 245. https://doi.org/10.3390/s22010245

Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B. and Rueckert, D. (2018) Attention U-Net: Learning where to look for the pancreas, arXiv:1804.03999. [Online]. Available: http://arxiv.org/abs/1804.03999

Prasad, P. J. R., Elle, O. J., Lindseth, F., Albregtsen, F. and Kumar, R. P. (2021) Modifying U-Net for small data set: a simplified U-Net version for liver parenchyma segmentation. In: *Proceedings of the SPIE* 11597, *Medical Imaging 2021: Computer Aided Diagnosis. SPIE Proceedings*, **11597**, Article 115971O. Society of Photo-Optical Instrumentation Engineers (SPIE). https://doi.org/10.1117/12.2582179

Priyadharsini, R. and Sharmila, T.S. (2019) Object detection in underwater acoustic images using edge based segmentation method. *Procedia Computer Science*, 165, 759–765. Available at: https://doi.org/10.1016/j.procs.2020.01.015.

Punn, N. S. and Agarwal, S. (2022) Modality Specific U-Net Variants for Biomedical Image Segmentation: a survey. *Artificial Intelligence Review*, **55**(7), 5845–5889. https://doi.org/10.1007/s10462-022-10152-1

Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: N. Navab, J. Hornegger, W. M. Wells and A. F. Frangi (eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science*, **9351**, 234–241. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

SANGUI, S., IQBAL, T., CHANDRA, P. C., GHOSH, S. K. AND GHOSH, A. (2023) 3D MRI Segmentation using U-Net Architecture for the detection of Brain Tumor. *Procedia Computer Science*, 218, 542–553. https://doi.org/10.1016/j.procs.2023.01.036

SIDDIQUE, N., PAHEDING, S., ELKIN, C. P. AND DEVABHAKTUNI, V. (2021) U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications. *IEEE Access*, **9**, 82031-82057, 2021, doi: 10.1109/ACCESS.2021.3086020.

SRIDHAR, B. (2020) A Quality Representation of Tumor in Breast Using Hybrid Model Watershed Transform and Markov Random Fields. In: *2020 International Conference on Computer Communication and Informatics (ICCCI)*, 1–5. Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/iccci48352.2020.9104154

WANG, X., CHANG, Z., ZHANG, Q., LI, C., MIAO, F. AND GAO, G. (2023) Prostate ultrasound image segmentation based on DSU-Net. *Biomedicines*, **11**(3), 646. https://doi.org/10.3390/biomedicines11030646

WANG, Z., LIU, Z., SONG, Y. AND ZHU, Y. (2019) Densely connected deep U-Net for abdominal multi-organ segmentation. In: *2022 IEEE International Conference on Image Processing (ICIP)*, 1415–1419, https://doi.org/10.1109/icip.2019.8803103

WANG, W., YU, K., HUGONOT, J., FUA, P. AND SALZMANN, M. (2019) Recurrent U-Net for Resource-Constrained Segmentation. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2142–2151. IEEE. https://doi.org/10.1109/iccv.2019.00223

WANG, R., ZHOU, H., FU, P., SHEN, H. AND BAI, Y. (2023) A Multiscale Attentional Unet Model for Automatic Segmentation in Medical Ultrasound Images. *Ultrasonic Imaging*, **45**(4), 159–174. https://doi.org/10.1177/01617346231169789

YANG, Z., SUN, X., YANG, Y. AND WU, X. (2024) MEDU-Net+: A Novel Improved U-Net Based on Multi-Scale Encoder-Decoder for Medical Image Segmentation. *KSII Transactions on Internet and Information Systems*, **18**(7), 1706-1725 https://doi.org/10.3837/tiis.2024.07.001

YU, H., HE, F. AND PAN, Y. (2019) A scalable region-based level set method using adaptive bilateral filter for noisy image segmentation. *Multimedia Tools and Applications*, **79**(9-10), 5743–5765. Available at: https://doi.org/10.1007/s11042-019-08493-1

ZHANG, J., LV, X., ZHANG, H. AND LIU, X. (2020) Aresu-net: Attention Residual U-Net for brain tumor segmentation. *Symmetry*, **12**(5), 721. https://doi.org/10.3390/sym12050721

ZHANG, Z., WU, C., COLEMAN, S. AND KERR, D. (2020) DENSE-INception U-Net for Medical Image Segmentation. *Computer Methods and Programs in Biomedicine*, 192, 105395. https://doi.org/10.1016/j.cmpb.2020.105395

Zhou, T., Ruan, S. and Canu, S. (2020) A Review: Deep learning for Medical Image Segmentation using Multi-modality Fusion. *Array*, 3–4, 100004. https://doi.org/10.1016/j.array.2019.100004

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N. and Liang, J. (2018) UNet++: A Nested U-Net Architecture for Medical Image Segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* 4th *International Workshop on DLMIA 2018 and 8th International Workshop on ML-CDS 2018, held in Conjunction with MICCAI 2018. Lecture Notes in Computer Science*, **11045**, 3–11. Springer, Cham. https://doi.org/10.1007/978-3-030-00889-5_1

Zhou, Z., Siddiquee, M. R., Tajbakhsh, N. and Liang, J. (2020) U-NET++: Redesigning Skip Connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, **39**(6), 1856–1867. https://doi.org/10.1109/tmi.2019.2959609