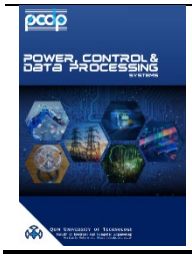
DOI: 10.30511/pcdp.2026.2090858.1081 - ISSN: 3060-8376 - Vol 3(4), 2026



POWER, CONTROL, AND DATA PROCESSING SYSTEMS

Available Online at: https://pcdp.qut.ac.ir/

Improvement of Breast Tumor Segmentation in Ultrasound Images by a New Effective Diversified Input U-Net Model

ARTICLE INFO

Article Type
Original Research

Authors
A. Seyed Hamid Khatami [1]
B. Nasser Mehrshad[2]*
C. Seyed Hamid Zahiri[3]

[1] Department of Electronics Engineering, Faculty of Electrical and Computer Engineering, University of Birjand, Birjand, Iran, h.khatami@birjand.ac.ir

2* ***Correspondence***
Department of Electronics Engineering, Faculty of Electrical and Computer Engineering, University of Birjand, Birjand, Iran, nmehrshad@birjand.ac.ir

[3] Department of Electronics Engineering, Faculty of Electrical and Computer Engineering, University of Birjand, Birjand, Iran, hzahiri@birjand.ac.ir

Article History
Received: June 7, 2026
Accepted: July 22, 2026
ePublished: December 01, 2026

ABSTRACT

This study aimed to improve breast tumor segmentation in ultrasound images using a modified U-Net architecture that reduces trainable parameters while enhancing segmentation accuracy and diagnostic performance. A lightweight modified U-Net with fewer trainable parameters was developed for breast ultrasound tumor segmentation using the BUSI and UDAIT datasets, and to compensate for the simplified structure and reduce overfitting risk, the input channels were diversified using complementary tumor and non-tumor region enhancement information, which improved both direct and indirect segmentation metrics while maintaining lower computational complexity than recent U-Net-based models. The model performance was evaluated using accuracy, precision, sensitivity, specificity, F1-score, Dice coefficient, and Intersection over Union (IoU). On the BUSI dataset, the proposed method achieved an accuracy of 0.99, precision of 0.95, sensitivity of 0.96, specificity of 0.99, F1-score of 0.95, Dice coefficient of 0.95, and IoU of 0.92; on the UDAIT dataset, it obtained an accuracy of 0.98, precision of 0.88, sensitivity of 0.96, specificity of 0.98, F1-score of 0.92, Dice coefficient of 0.92, and IoU of 0.85. Compared with previous approaches, the proposed method achieved superior segmentation performance, particularly in Dice coefficient and F1-score, while maintaining lower computational complexity due to fewer trainable parameters. In conclusion, the proposed modified U-Net model improved breast tumor segmentation in ultrasound imaging while maintaining lower computational complexity than newer U-Net variants, and its high accuracy and Dice scores indicate its potential as an efficient decision-support tool for breast cancer diagnosis.

Keywords: Breast Tumor Segmentation; Diversified Input U-Net; Image processing; Ultrasound.

1 Introduction

Breast cancer poses a significant risk to women's health. Diagnosing the pathogenic factors in breast cancer and choosing the appropriate treatment options can be complex and time-consuming [1]. However, early detection is crucial, as the survival rate for early-stage breast cancer exceeds 90% [2].

Ultrasound imaging is a quick, safe, accessible, and cost-effective diagnostic tool for breast cancer that uses sound waves to produce images of the breast. This method allows for real-time imaging and can accommodate multiple exams within a short period. It is suitable for women of all ages, including those who are breastfeeding. However, this technique has certain limitations, such as low image resolution, blurriness, and unclear tissue boundaries[3]. Additionally, the different shapes of breast lesions can make clinical diagnosis more challenging for physicians, especially in high-pressure situations with heavy workloads, which might lead to incorrect diagnoses. As a result, there is an increasing need for computer-aided diagnostic (CAD) systems to assist physicians. CAD technology improves the accuracy of breast examinations and supports early screening for breast cancer [4].

To address the challenges of breast cancer diagnosis, deep learning models enhance the accuracy of diagnoses by automatically analyzing cancerous lesions, resulting in more reliable assessments. This improved accuracy and efficiency in diagnosis facilitate better-informed treatment choices. In recent years, deep learning techniques have advanced significantly in the field of medical image analysis, leading to new possibilities for the automatic segmentation of breast ultrasound images [5, 6]. Automated segmentation allows radiologists and oncologists to evaluate the size, shape, and spatial characteristics of tumors, which aids in treatment planning and monitoring disease progression [7].

Computer-aided diagnostic (CAD) systems utilize techniques such as image feature extraction, pattern recognition, and tissue structural analysis, along with algorithms, to segment breast lesions. This process aids in identifying and localizing these lesion regions, ultimately enhancing the accuracy and objectivity of clinical diagnoses when combined with the physician's assessment [8, 9]. In traditional methods, gradients and gray-level features are used to delineate lesion areas. These methods are based on a local comparison of intensity values and are implemented using Finite Impulse Response (FIR) filters. These approaches do not take into account the overall quality of the image or the expert intuition regarding it. This limitation also applies to techniques that utilize deep convolutional neural networks, as they function similarly to FIR filters. These methods can be effective in simpler situations, but they often struggle with more complex scenarios and different image generalizations. While deep learning methods aim to incorporate higher-order features and have made significant strides in handling complex scenes, noisy images, and tasks that demand high levels of detail, they still fall short when it comes to applying expert intuition to understand the entirety of an image [10-13].

While the U-Net architecture has demonstrated strong performance in medical segmentation, its standard model has limitations, particularly in breast tumor segmentation. As a result, various modified architectures have been developed to enhance performance by incorporating attention mechanisms and multi-scale structures. However, these improvements also lead to increased computational complexity due to the rise in the number of parameters [14, 15].

Recently, several hybrid and attention-driven architectures have been proposed to further enhance medical image segmentation. For instance, attention-guided Swin Transformer networks have been effectively utilized for 3D brain tumor segmentation, leveraging global contextual modeling to improve interpretability [16]. In the domain of breast ultrasound, the synergistic role of data augmentation and attention mechanisms has been systematically explored within U-Net and Conv-NeXt frameworks, demonstrating that strategic preprocessing and feature recalibration can significantly boost segmentation performance [17]. Moreover, vision transformers combined with axial state-space modeling have shown promise in capturing long-range dependencies for 3D tumor delineation [18]. Additionally, GAN-boosted U-Net architectures have been proposed to enhance breast tumor segmentation by generating synthetic features that improve both performance and model interpretability [19]. While these deep learning approaches achieve impressive results, they require extensive annotated data, heavy computational resources, and long training times. In contrast, our proposed method achieves competitive accuracy by intelligently diversifying the input channels of a lightweight U-Net, thereby avoiding the complexities and data-hungry nature of transformer-based or GAN-based models.

While the developed versions of U-Net have demonstrated effective performance, some research has highlighted structural challenges associated with this architecture. One significant issue is the reduction of spatial accuracy in the deeper layers of the network, which is crucial for accurately segmenting breast lesions in ultrasound images [20]. Additionally, increasing the number of convolutional masks in these layers can further complicate this problem.

The developed U-Net models aim to enhance the efficiency and accuracy of segmentation by increasing parameters and structural complexity. However, this increase in parameters can lead to significant challenges during the training phase. When there are many learning parameters, a substantial amount of labeled data is required to mitigate the risk of overfitting. Additionally, this complex structure can pose a challenge due to the high computational demands during the testing phase. In this paper, we simplify the standard U-Net architecture to address these structural challenges and reduce the likelihood of overfitting. By purposefully diversifying the input for this simplified model, we have significantly improved the method's efficiency compared to existing

complex structures. In summary, our main contributions are as follows:

- Reducing the parameters of the standard U-Net model while deliberately increasing the input channels aims to improve the segmentation of tumor regions in breast ultrasound images.
- Using complementary tumor/non-tumor region enhancement inputs to improve direct and indirect tumor segmentation evaluation indices simultaneously.
- Achieving greater computational efficiency than newer versions of the U-Net-based architectures, despite the increased number of input channels, by utilizing the lightweight U-Net architecture during training.

The remaining sections of the manuscript are structured as follows: The second section reviews the literature. The third section describes the materials and methods, which include the dataset, preprocessing techniques, input diversification strategy, U-Net structure simplification, proposed algorithm, implementation settings, and evaluation criteria. The final two sections present the results of the method, followed by a discussion and conclusion.

2 Literature review

Breast ultrasound imaging is valued for its non-invasive, accessible, affordable, and quick procedures. This makes it an important tool for diagnosing and monitoring breast lesions within a patient-centered healthcare system [21]. However, the lower resolution of breast ultrasound images, compared to some other imaging methods, presents challenges. Issues like tissue clutter, blurred boundaries, and large variations in the shape and size of lesions can make their accurate differentiation more difficult [22].

In the automatic detection of breast lesions in ultrasound images and the monitoring of their treatment, the first and most crucial step is lesion segmentation. This process helps determine the size, shape, and location of the lesion. With the advancement of deep learning techniques, a fundamental model for medical image segmentation known as U-Net was introduced in 2015 [23]. This straightforward yet effective architecture features an encoder-decoder pathway that adds depth to the network while preserving important image details through the use of “jump connections” at relevant levels. Because of these characteristics, U-Net has become a foundational model for various breast tumor segmentation applications in ultrasound imaging [24].

In subsequent research, modifications were made to the U-Net architecture to enhance the accuracy and stability of the results. The U-Net++ model was developed, which redesigned the connections between the encoder and the decoder to improve the combination of features at multiple scales [25]. In this architecture, instead of directly linking each level of the encoder to the corresponding level of the decoder, a set of intermediate nodes was introduced. These nodes provide multiple pathways for transmitting information, enabling the better preservation of information across different levels of network depth. As a result, this leads to more accurate segmentation. However, a drawback of this architecture is its increased structural complexity compared to the basic U-Net. It features a higher number of parameters and requires a longer training time. Additionally, in some cases, the magnification of fine details, such as lesions, may not be adequately addressed [26].

The U-Net3+ architecture builds upon the development of the U-Net framework. This architecture is specifically designed to incorporate full-scale skip connections and deep supervision. In U-Net3+, each layer in the decoder section utilizes not only the corresponding encoder layer but also all encoder levels and some higher reconstruction levels. This means that high-level and low-level features are combined more extensively, allowing the model to consider both detailed and general information [27]. Studies have reported improved performance for U-Net3+ compared to both U-Net and U-Net++. It achieves a better Dice coefficient with fewer parameters than U-Net++, and the extensive combination of multi-scale features enhances the detection of boundaries and objects of varying sizes. This advantage is particularly valuable for medical images that contain complex lesions. However, despite the reduction in the number of parameters noted in some studies, the overall structure of U-Net3+ remains relatively complex. This complexity can lead to increased training time and higher resource requirements. Additionally, some research has indicated that in cases with a small number of samples or tiny objects, U-Net3+ may still struggle, particularly when dealing with ultrasound images that have blurred boundaries and dense textures, unless further optimization is applied [28].

In another study, adding Attention Gates to the U-Net architecture allows the model to emphasize relevant parts of an image, such as tumor regions, while minimizing the focus on irrelevant areas. This approach enhances the sensitivity and accuracy of segmentation in images with complex structures without necessitating significant structural changes. Additionally, it typically incurs a relatively small computational cost compared to integrating larger convolutional layers. However, under conditions with minimal data, substantial improvements may not always be observed, even though this method does not significantly increase the parameter burden. Furthermore, attention mechanisms can sometimes become overly focused, leading to poorer interpretation of fine boundaries. In practice, while there is often an increase in metrics such as Dice and Recall, improvements in Intersection over Union (IoU) or reductions in false positives may vary depending on the training settings [29].

Researchers also integrated the concepts of transformers, which utilize holistic self-attention, with the U-Net architecture, leading to the development of models such as Trans U-Net. They also explored hierarchical transformer

architectures, such as Swin-Transformer, arranged in a U-shaped form (Swin U-Net). These approaches enhance the model's ability to capture global context and long-range relationships within images, leading to improved Dice scores and more stable results. However, it's important to note that transformers typically have a higher number of parameters and greater computational costs. They also require more data for effective training, and without sufficient data, they may struggle with accurate localization. When applying these models to noisy images, such as those from ultrasound, additional strategies like pre-processing, regularization, and augmentation are necessary [30].

CA-Net integrates three types of attention modules-spatial, channel, and scale-into the network. This combination allows the model to focus on the most important areas of an image, recalibrate channel responses, and address different scales of damage. In its experiments, CA-Net demonstrates a significant improvement in Dice scores compared to the baseline U-Net. However, it is important to note that this approach may increase computational complexity and might not deliver substantial enhancements when working with minimal data sets [31].

Swin U-Net is also based on a transformer architecture and specifically uses Swin transformer blocks to model long-range dependencies in an image more effectively. This method is generally superior to some convolutional or hybrid approaches. The main advantage of using Swin U-Net is its improved ability to understand the whole of an image and connect distant regions. However, this technique also has some disadvantages: it requires more data and computational resources and has a higher risk of overfitting, especially when limited data is available [32].

The nnU-Net framework takes a unique approach compared to architectures like U-Net3+ or those that incorporate attention mechanisms. nnU-Net, which stands for "no-new-Net," focuses on utilizing the basic U-Net architecture while automating parameter tuning, preprocessing, training, and inference for each dataset. Instead of developing a new network structure, nnU-Net aims to streamline the application settings for different medical data. This method is particularly advantageous when there is significant variability in data types, such as image modality, lesion size, resolution, and sources, making manual model tuning challenging and prone to errors. However, a limitation of this approach is that the underlying network architecture does not change compared to methods that enhance performance by adding specific modules, like attention or transformers. As a result, nnU-Net may struggle in scenarios that require modeling very long-range dependencies or intricate textures. However, the reference notes that nnU-Net has achieved top rankings on multiple medical image segmentation challenges without the need for manual intervention [33].

The basic model currently used for segmenting breast ultrasound images lacks the structural intelligence needed to adapt effectively to this specific task. Additionally, the unique characteristics of these images have not been utilized to develop preprocessing layers or parallel preprocessing paths. If we aim to enhance a basic model by adding complexity to achieve better results in this specific application, we must focus on two key aspects: simplifying the task itself and optimizing the tools required to perform the work. Ignoring the unique challenges of breast tumor image segmentation in ultrasound images and merely increasing the complexity of a general segmentation model will not lead to optimal performance or effective solutions.

In this study, we aim to achieve both accurate and efficient tumor lesion segmentation in breast ultrasound images. Our approach involves simplifying the model structure by reducing the number of parameters while also enriching the inputs by increasing input channels or utilizing parallel preprocessing. It is important to note that previous methods aimed at enhancing and modifying the U-Net architecture through an increase in parameters and complexity have not resulted in a significant improvement in evaluation metrics. Additionally, given the patient-centered focus of the healthcare system, these enhancements often do not justify the added complexity and computational burden. While many U-Net variants achieve higher accuracy on specific benchmarks, their performance gains come at the cost of significantly increased parameter counts and training data requirements. Notably, the nnU-Net framework demonstrated that a well-configured standard U-Net often outperforms more complex hand-crafted architectures across multiple medical segmentation tasks when properly optimized. This observation, also discussed in recent systematic reviews, underscores that architectural complexity does not universally guarantee better performance, particularly in small-dataset scenarios like breast ultrasound segmentation.

3 Materials and methods

In this section, we outline our research methodology. We start by describing the dataset, then provide a detailed explanation of our preprocessing techniques and input diversity strategy. Next, we introduce the proposed U-Net model, which is designed to be minimally complex. We present the overall tumor segmentation algorithm, followed by a description of the experimental setup and the evaluation criteria.

3.1 Dataset

In this study, two datasets were used to evaluate the proposed method for detecting and segmenting breast cancer tumors in ultrasound images. First, known as BUSI, is a reliable and comprehensive source for detecting and segmenting breast tumors, which is specifically designed for developing image processing and machine learning algorithms.

Table 1: Distribution of sample sizes for datasets

	Benign	Malignant	Format	Total
BUSI	487	210	PNG	697
UDAIT	110	53	PNG	163

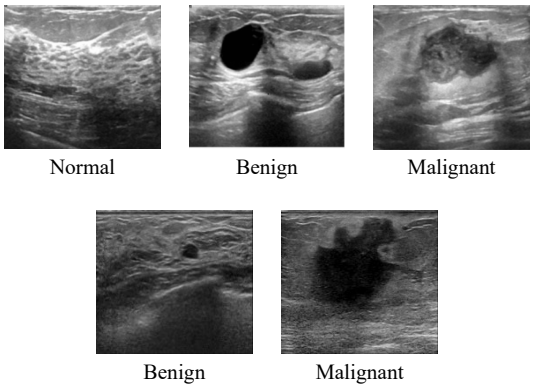


Figure 1: Sample images of the BUSI (top) and UDAIT (bottom) Datasets [34]

This dataset comprises a diverse set of images accurately labeled by radiologists into three categories: benign, malignant, and normal. In addition, ground-truth images for tumor segmentation are available from different medical centers, collected to simulate real-world conditions by incorporating natural noise and substantial tumor variability.
In this paper, we have used benign and malignant images from this dataset. The second dataset, UDAIT, was recorded using a Siemens ACUSON device at the UDAIT diagnostic center in Spain. It contains both benign and malignant images. Table 1 lists the number of samples in each dataset, while Figure 1 presents sample images from these datasets.

3.2 Pre-processing and Input Diversity Strategy

In breast ultrasound images, accurately detecting tumor areas requires careful attention to the darker regions of the image, while correctly identifying normal tissue involves focusing on the brighter areas. Enhancing contrast in dark areas can aid in tumor detection, whereas reducing contrast there or increasing it in brighter regions can help identify normal tissue. This creates a trade-off between focusing on dark regions and bright regions. To address this, we implement a parallel preprocessing approach. The key innovation of this research is the use of a 7-channel input instead of a single-channel grayscale image. Each channel is a nonlinear transformation of the original image, which increases spectral diversity, highlights lesion boundaries and textures, and enables the model to better recognize different brightness, texture, and intensity patterns.
Transformations such as I^2, I^3, and I^4 amplify higher intensity values while compressing lower ones. As a result, they suppress background variations in darker regions (which often include tumor cores in ultrasound) and push these areas toward relative uniformity, while enhancing contrast in brighter tissues. In contrast, root functions such as $\sqrt{I}$ and $\sqrt[3]{I}$ expand lower intensity ranges and compress higher ones. This behavior enhances contrast in darker regions, making hypoechoic tumor areas more distinguishable from surrounding tissue, while simultaneously reducing intensity differences in brighter non-tumorous regions and driving them toward greater uniformity.
Therefore, root transformations $\sqrt{I}$ and $\sqrt[3]{I}$ are particularly effective at highlighting tumor regions while homogenizing non-tumorous bright areas. Meanwhile, higher-order power transformations (I^2, I^3, I^4) emphasize bright structures and help the model better recognize peripheral and side areas of the lesion that typically have low contrast with surrounding tissue. The combination of these complementary transformations enhances the model's ability to represent lesions accurately and enables it to extract more diverse and discriminative information from pixel intensities. In Figure 2, a sample image of these seven channels is shown. Although some channels may appear visually similar, each reveals distinct information about texture features and pixel intensity distributions, which is essential for maintaining the balance of the proposed method.

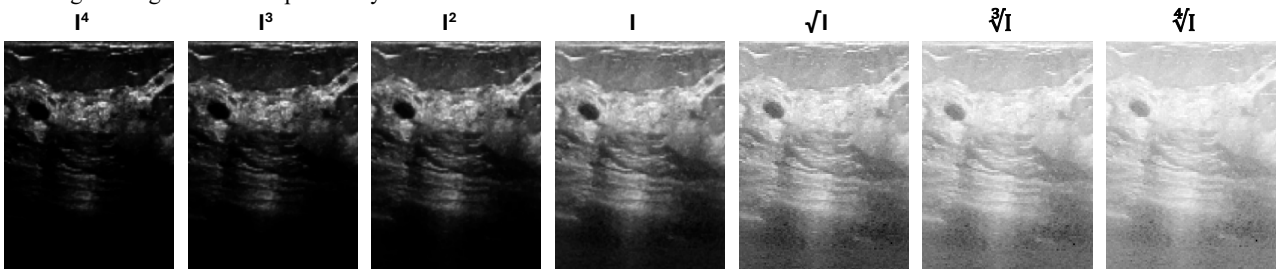


Figure 2: Sample image, seven channels

Our contribution is not merely a set of preprocessing steps, but a principled framework that couples aggressive architectural simplification with spectrally diverse input representations. While existing lightweight U-Net variants reduce parameters at the cost of accuracy, our diversified input strategy maintains and even improves segmentation performance by providing the network with multiple views of the same intensity distribution. This differs fundamentally from handcrafted feature engineering, as we do not extract or select features; rather, we enrich the input space so that the network can implicitly learn the most discriminative intensity relationships during training. The seven transformations were selected based on ultrasound physics: root functions ($\sqrt{I}$, $\sqrt[3]{I}$, $\sqrt[4]{I}$) expand low-intensity ranges to highlight hypoechoic tumor regions, while power functions (I^2, I^3, I^4) compress low intensities and emphasize high-intensity boundaries, and the original intensity

(I) preserves the raw clinical appearance. This balanced coverage of the dynamic range is the key to our method's success.

3.3 Proposed U-Net model with the least complexity

As previously mentioned, this study aims to simplify breast tumor ultrasound image segmentation by diversifying the input to the deep network in the first stage. In the next stage, we propose a streamlined U-Net structure that uses the fewest parameters and minimal coding and decoding steps to achieve effective segmentation of these images. Although it is possible to improve this U-Net network with more advanced architectures, such as U-Net++ and U-Net3+, as shown in Figure 3, our main goal is to demonstrate that diversifying the input can boost segmentation results. Therefore, we have decided not to include these structural enhancements in this study and will address them in future work.

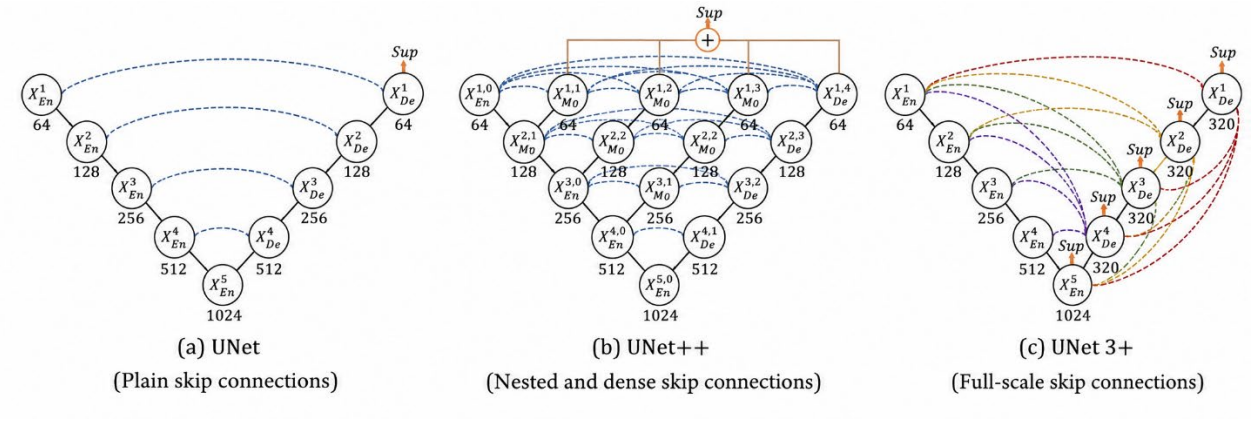


Figure 3: Comparison of U-Net, U-Net++, and U-Net 3+ (c) [27]

In this study, we have used a simplified version of the basic U-Net model. U-Net is a fully connected convolutional network with a symmetric encoder–decoder structure that enables accurate learning of spatial structures by utilizing a contracting path to extract high-level features and an expanding path to reconstruct the segmentation map.

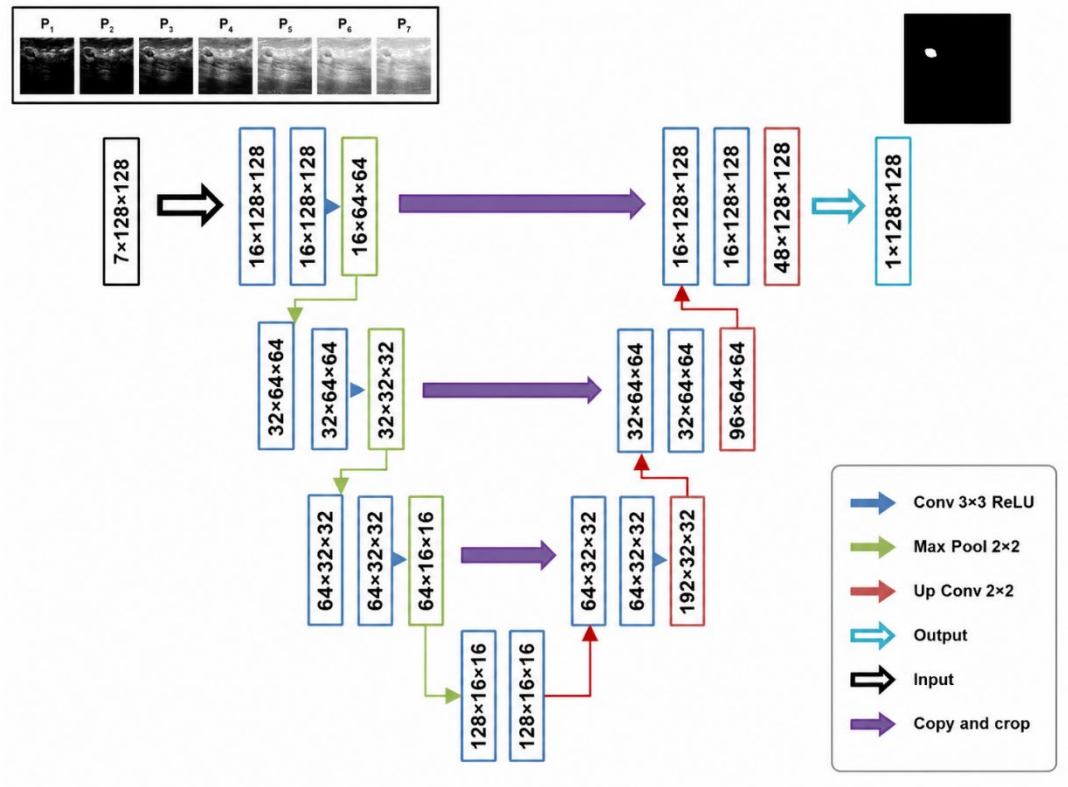


Figure 4: U-Net Architecture

In this study, our input is a seven-channel input with a size of 128×128 pixels, which is fed into several convolution layers with 16, 32, 64, and 128 filters in the Encoder or Down Sampling path.

After each convolution layer, a Max-Pooling operation is performed to reduce spatial dimensions, and important features are extracted. After reaching the lowest level or bottleneck, the network is a two-layer convolution with 128 filters. In the Decoder or Up-Sampling path, the image size is increased, and spatial information is recovered to preserve details.

Table 2: Layers in the U-Net model

Layer Name	Layer Type	Input Shape	Output Shape	Param
input	Input	None	(128, 128, 7)	0
conv1_1	Conv2D	(128, 128, 7)	(128, 128, 16)	1024
conv1_2	Conv2D	(128, 128, 16)	(128, 128, 16)	2320
pool1	MaxPooling2D	(128, 128, 16)	(64, 64, 16)	0
conv2_1	Conv2D	(64, 64, 16)	(64, 64, 32)	4640
conv2_2	Conv2D	(64, 64, 32)	(64, 64, 32)	9248
pool2	MaxPooling2D	(64, 64, 32)	(32, 32, 32)	0
conv3_1	Conv2D	(32, 32, 32)	(32, 32, 64)	18496
conv3_2	Conv2D	(32, 32, 64)	(32, 32, 64)	36928
pool3	MaxPooling2D	(32, 32, 64)	(16, 16, 64)	0
bottleneck1	Conv2D	(16, 16, 64)	(16, 16, 128)	73856
bottleneck2	Conv2D	(16, 16, 128)	(16, 16, 128)	147584
up3	UpSampling2D	(16, 16, 128)	(32, 32, 128)	0
concat3	Concatenate	(32, 32, 128) +(32, 32, 64)	(32, 32, 192)	0
conv4_1	Conv2D	(32, 32, 192)	(32, 32, 64)	110656
conv4_2	Conv2D	(32, 32, 64)	(32, 32, 64)	36928
up2	UpSampling2D	(32, 32, 64)	(64, 64, 64)	0
concat2	Concatenate	(64, 64, 64) +(64, 64, 32)	(64, 64, 96)	0
conv5_1	Conv2D	(64, 64, 96)	(64, 64, 32)	27680
conv5_2	Conv2D	(64, 64, 32)	(64, 64, 32)	9248
up1	UpSampling2D	(64, 64, 32)	(128, 128, 32)	0
concat1	Concatenate	(128, 128, 32) +(128, 128, 16)	(128, 128, 48)	0
conv6_1	Conv2D	(128, 128, 48)	(128, 128, 16)	6928
conv6_2	Conv2D	(128, 128, 16)	(128, 128, 16)	2320
output	Conv2D	(128, 128, 16)	(128, 128, 1)	17

After each Up-Sampling and merging operation, two convolution layers are applied with fewer filters than in the previous stage. Finally, a 1×1 convolution layer is used, and the Sigmoid activation function produces the final output. This output is a binary image with a size of 128×128, having one channel, and predicts the tumor area in the ultrasound image. In Figure 4 and Table 2, you see the steps of a U-Net network. Also in Table 3, see a comparison between the number of parameters per channel and the computational complexity of each channel in the proposed U-Net model. As is clear, increasing the number of channels does not lead to a significant increase in the number of parameters, and no significant increase in computational complexity is observed. At the same time, this increase in the number of channels has been purposeful and has increased the accuracy of the model.

The reason for the precise selection of these seven functions and the avoidance of using fewer or more is the result of an experimental and theoretical analysis of the effect of each transform on the ultrasound tissue. Although using only 3 or 4 channels (e.g., only I, I^2, $\sqrt{I}$), it covers part of the intensity information; it cannot provide the necessary balance between very bright areas (such as the lesion border) and very dark areas. In particular, high powers such as I^3 and I^4 are important for enhancing high-intensity areas and increasing contrast at the blurred borders of the lesion.

Table 3: Comparison of the computational complexity of our proposed U-Net method with a different number of channels

Number of channels	Parameters	FLOPs
1	0.487009M	0.487009G
3	0.487297M	1.461891G
5	0.487586M	2.437925G
7	0.487873M	3.415111G

In contrast, high roots such as $\sqrt[3]{I}$ and $\sqrt[4]{I}$ contribute to better representation of dark and low-contrast areas, which is of particular importance in ultrasound, since lesions often appear in a background of low-intensity noise. The use of 5, 6, 8, 9, 10, and 11 channels has also been investigated, but the results showed that either certain information (e.g., very high or very low intensities) is omitted or excessive overlap between channels is created, leading to reduced distinction in the feature space.

The careful selection of seven channels is designed with balanced coverage of the full spectrum of image intensities (from darkest to brightest). This selection is not only computationally lighter than methods based on complex feature extraction, but also increases the variety of information fed to the model, without creating unhelpful redundancy in the channels.

This combination allows the model to more richly understand fine tissue structures, lesion boundaries, and the distinction between lesion and background. Therefore, this choice was made not only based on trial and error but also with a deep

understanding of the nature of ultrasound imaging and the mathematical properties of transformations

The important point in this implementation is the use of 7-channel input images. This design increases the spectral diversity of the data and the richness of the input features, which, in turn, improves the model's performance in detecting tumor boundaries and areas more accurately.The model's output also generates a binary probability map (tumor/non-tumor tissue) using a 1×1 convolution layer with a sigmoid activation, reflecting the lowest complexity in the U-Net network's structure.

We have not used the developed versions of the U-Net model, which have much more complex and time-consuming operations, and this simplicity is one of the advantages of our proposed method and the result of the increased diversity in the input data of the U-Net network.

It also reaches convergence in a smaller number of Epochs. In the basic U-Net model, fluctuations in validation can indicate overfitting. While in the proposed U-Net model, the gap between the training and validation curves is reduced, and the validation graph is more stable. Also, in the training process of the proposed U-Net model, higher accuracy was achieved in a lower number of epochs. For example, in an epoch of 15, it reached an accuracy of 0.97, which is 0.96 for the basic U-Net model.

In contrast, the proposed model, by utilizing an optimized architecture and lightweight design, has managed to provide significant efficiency in terms of memory consumption with a 97.2% reduction in the number of parameters compared to the base U-Net. These results indicate that the proposed model, while maintaining the ability to extract effective features, has much lower complexity and is well-suited for environments with hardware constraints.

The proposed U-Net architecture was designed with a minimal number of trainable parameters to reduce overfitting risk. Detailed layer configurations and theoretical computational complexity (FLOPs) are presented in Table 2 and Table 3, respectively. The experimental evaluation of different input channel configurations and the training dynamics are reported in Section 4.

3.4 Proposed tumor segmentation algorithm and experimental settings

In this study, we aim to accurately segment tumors in breast ultrasound images and use an improved U-Net model for this purpose. Our input data are from the public BUSI and UDAIT datasets, which consist of grayscale ultrasound images with a tumor region mask manually determined by expert clinicians. Figure 5 shows the steps of our proposed method, which are described in detail below.

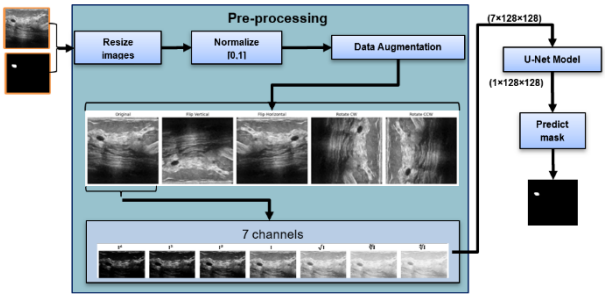


Figure 5: Different steps of our proposed method

In this algorithm (as you can see in Figure 5), first, all images are resized to 128×128, and the pixel values are normalized between 0 and 1. Using Equation 1, the images are normalized.

$$I_{norm}(x,y) = \frac{I(x,y) - I_{min}}{I_{max} - I_{min}} \quad \text{Eq 1}$$

$I(x,y)$ is the pixel matrix of the original image, I_{min} is the minimum pixel value in the image, and I_{max} is the maximum pixel value in the image. To increase the accuracy and generalization, we implement the data augmentation technique and, as shown in Figure 5, 6 more images are generated from each original image, including a vertically inverted image, a horizontally inverted image, and a 90° clockwise and 90° counterclockwise rotation.

Table 4 shows the number of images in the dataset before and after the data augmentation technique. Then, from each of these images, 6 more images with different brightness intensities are generated, as shown in Figure 2, and they constitute the 7 input channels of the proposed U-Net model, which is the key innovation of our algorithm and provides important contextual information to the model. Then, using the U-Net model and its encoder and decoder structures, the tumor area is predicted.

The U-Net model used was designed with a conventional architecture, including an encoding path and decoding path, and a binary cross-entropy error function and Adam optimizer were used to train it. After training, the model's performance was evaluated, and the results were calculated based on criteria such as accuracy, sensitivity, specificity, F1-score, IoU, and Dice coefficient.

Table 4. Model training settings for BUSI and UDAIT datasets

Dataset	Input image size	Train	Test	Total samples (Without augmentation)	Total samples (With augmentation)	Total samples after conversion to 7 channels
BUSI	7×128×128	%80	%20	697	3485	24395
UDAIT				163	815	5705



Figure 6: Sample of Data augmentation technique output images

To prevent data leakage, we first performed a stratified 80/20 split of the original dataset into training and testing subsets. All data augmentation operations, including vertical/horizontal flips and 90° rotations, were applied only to the training images after the split. The augmented training set was then used to generate the seven-channel inputs. The testing set remained untouched and was used solely for final evaluation. This guarantees that no augmented version of any test image appears in the training set, ensuring unbiased performance assessment.

The experiments were conducted in the Google Colab environment using hardware including a 2.20 GHz dual-core Intel Xeon processor, 12 GB of RAM, and an NVIDIA Tesla T4 graphics card with 16 GB of memory.

The base operating system of this environment was a version of Ubuntu 18.04, which runs on the Linux platform. Also, the development and execution of the model were carried out using software such as Python 3, TensorFlow, OpenCV, scikit-image, and matplotlib.

All models were trained using the Adam optimizer with a fixed learning rate of 1×10^{-4}, a batch size of 8, and a binary cross-entropy loss function. A random seed of 42 was set for all experiments to ensure reproducibility. Early stopping with a patience of 5 epochs was applied based on validation loss to prevent overfitting (Table 5).

Table 5. Training hyperparameters used in all experiments

Hyperparameter	Value / Setting
Optimizer	Adam
Learning rate	1×10^{-4}
Batch size	8
Number of epochs	20
Loss function	Binary Cross-Entropy (BCE)
Random seed	42
Learning rate decay	None (constant)
Validation split	20% of training set (early stopping)

3.5 Evaluation criteria

The effectiveness of a tumor segmentation algorithm is evaluated using four main criteria: True positive (TP) and true negative (TN) represent the number of tumor and non-tumor pixels that are correctly identified, respectively, while false positive (FP) and false negative (FN) represent the number of non-tumor and tumor pixels that are misclassified, respectively.

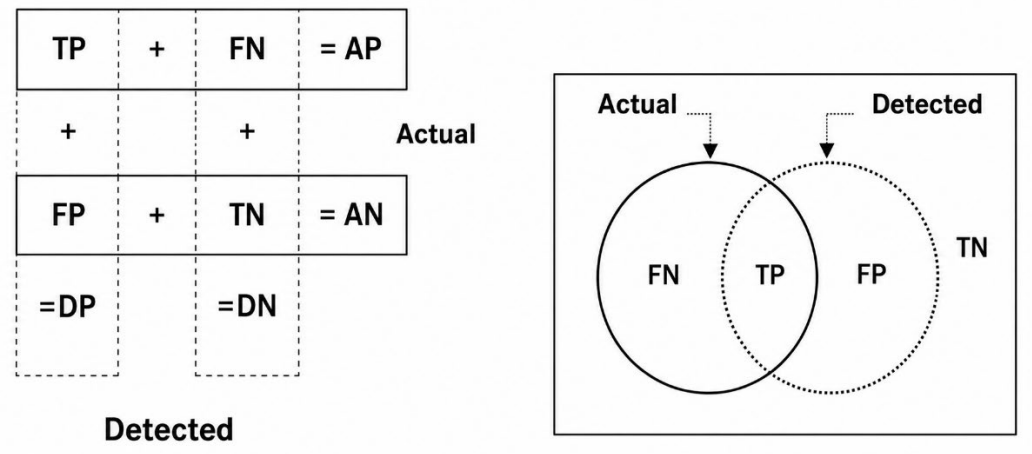


Figure 7: Complete description of the tumor location predictor and detected condition[35]

Additionally, AP and AN represent the actual number of tumor and non-tumor pixels, respectively. In contrast, DP and DN represent the number of pixels identified as tumor and non-tumor, respectively (as shown in Figure 7).

To evaluate the model, we use common evaluation metrics. These metrics include: Accuracy, Precision, Sensitivity, Specificity, F1-score, Dice scores, and IoU. Here, PPV is the positive predictive value. These evaluation criteria are shown in Equations 2 to 8.

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} = \frac{TP+TN}{AP+AN} = \frac{TP+TN}{DP+DN} \quad \text{Eq 2}$$

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{DP} \quad \text{Eq 3}$$

$$Sensitivity = \frac{TP}{TP + FN} = \frac{TP}{AP} \quad \text{Eq 4}$$

$$Specificity = \frac{TN}{AN} = \frac{TN}{TN + FP} = \frac{TN}{AN} \quad \text{Eq 5}$$

$$F1\text{-}score = 2 \times \frac{PPV \times SN}{PPV + SN} = \frac{2 \times TP}{2 \times TP + FP + FN} = \frac{2 \times TP}{DP + AP} \quad \text{Eq 6}$$

$$IoU = \frac{|A \cap B|}{|A \cup B|} = \frac{Dice}{2 - Dice} = \frac{TP}{TP + FP + FN} = \frac{TP}{DP + FN} \quad \text{Eq7}$$

$$Dice = \frac{|A \cap B| \times 2}{|A| + |B|} = \frac{2 \times IoU}{1 + IoU} = \frac{TP + TP}{TP + TP + FP + FN} = \frac{2 \times TP}{2 \times TP + FP + FN} = \frac{2 \times TP}{DP + AP} \quad \text{Eq 8}$$

4 Results and Conclusions

4.1 Model Convergence and Parameter Efficiency Analysis

To validate the efficiency of our simplified architecture, we first analyzed the impact of the number of input channels on segmentation performance. As shown in Figure 8, increasing the number of channels from 1 to 7 progressively improves the Dice coefficient, with the peak performance achieved at 7 channels. Beyond 7 channels (up to 11), the performance plateaus and slightly declines due to feature redundancy, confirming that the proposed seven-channel diversification provides the optimal trade-off between information richness and generalization.

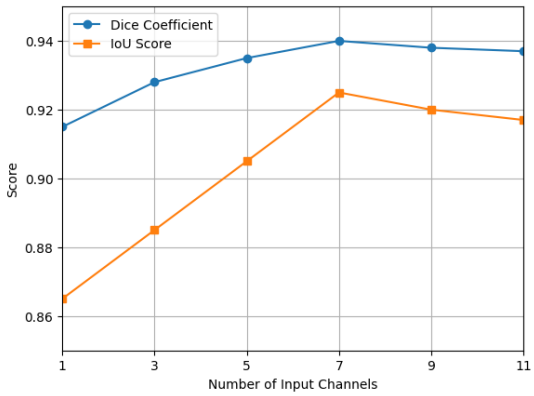


Figure 8: The effect of the number of input channels

Figure 9 compares the training and validation accuracy curves of the standard U-Net and our proposed lightweight model. While both models eventually reach high training accuracy (~99%), our proposed model exhibits a smoother and more stable convergence trajectory with significantly smaller fluctuations in validation accuracy. Notably, the proposed model achieves a validation accuracy of 0.97 by epoch 15, whereas the standard U-Net reaches only 0.96 at the same epoch, with visible oscillations. The reduced gap between training and validation curves in our method indicates better generalization and lower overfitting, which is attributed to the 97.2% reduction in trainable parameters.

Table 6 provides a comprehensive comparison of the parameter counts between our method and several state-of-the-art U-Net variants. While architectures such as Trans U-Net and Swin U-Net increase parameters by up to 480% to achieve marginal gains, our method contains only 0.487 million parameters, a reduction of 97.2% compared with the standard U-Net (17.27M). This drastic simplification, compensated by the diversified input strategy, forms the cornerstone of our method's computational efficiency.

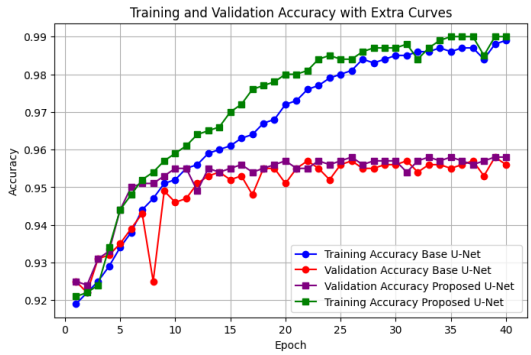


Figure 9: Comparison of training and validation accuracy graphs in both the base U-Net model and the proposed U-Net model

Table 6: Comparison of Model Parameters with Baseline U-Net

Model	Number of parameters	~%Parameters (VS. baseline U-Net)
U-Net	17.27M	1× (baseline)
U-Net++	9.16M	↓ 47%
U-Net3+	30-35M	↑ 73.7–102.6%
Attention U-Net	19-21M	↑ 10–21.6%
Trans U-Net	93.19M	↑ 439.6%
Swin U-Net	60-100M	↑ 247-479%
NN U-Net	30.58M	↑ 77.1%
CA U-Net	30.18M	↑ 74.8%
Proposed method	**0.487M**	↓ **97.2%**

4.2 Results

This study investigates the performance of a simplified version of the basic U-Net model, enhanced by an input enrichment through a seven-channel input technique and data augmentation. The model's effectiveness in delineating suspicious areas in breast ultrasound images has been evaluated through both quantitative and qualitative methods. Figure 10 illustrates an example of the predicted masks that represent tumor areas. These masks demonstrate a high level of agreement with the ground truth masks, as assessed using similarity and overlap criteria.

The results indicate that the proposed input enrichment algorithm successfully identifies and separates tumor areas with high accuracy, even with the simplification of the U-Net network. Furthermore, Tables 7 and 8 compare the performance of the proposed method with a wide range of

state-of-the-art architectures, mainly based on different versions of the U-Net and its extensions. The comparison is based on the metrics of Accuracy, Precision, Sensitivity, Specificity, and F1 score on two datasets, BUSI and UDAIT.

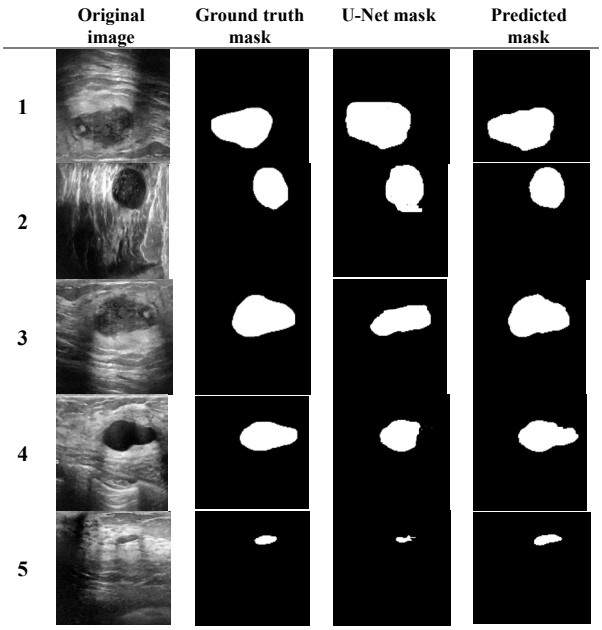


Figure 10: Sample output images of the proposed algorithm

On the BUSI dataset, the proposed method achieves an accuracy of 0.99, a sensitivity of 0.96, a specificity of 0.99, a positive precision of 0.95, and an F1 score of 0.95. These results are superior to several well-known architectures such as U-Net++, U-Net3+, SK U-Net, CAU-Net, and Attention Dense U-Net.

For example, although CAU-Net and Attention Dense U-Net report F1 scores of 0.87 and 0.88, respectively, they are still lower than the value of 0.95 of the proposed method.

Also, compared to Asym U-Net and PDF U-Net, which report lower sensitivities (0.78 and 0.62, respectively) in some experiments, the proposed method, with a sensitivity of 0.96, shows a much higher ability in lesion detection. This improvement in sensitivity is of particular importance in medical applications, as it means a reduced false negative rate and thus a reduced chance of missing tumor areas.

It is noteworthy that many of the compared methods use advanced architectural modifications such as dense connections, residual blocks, attention mechanisms, or multi-scale feature aggregation.

In contrast, the proposed method outperforms, or at least matches, the best methods without using complex attention-based models or highly parameterized U-Net variants. This demonstrates the efficiency and robustness of the proposed framework in accurately segmenting breast lesions while maintaining architectural simplicity.

Table 7: Comparison of the proposed method with other architectures based on Accuracy, Sensitivity, Specificity, and F1-Score

Dataset	Ref	Architecture	Accuracy	Precision	Sensitivity	Specificity	F1
UDAIT	[36]	Asym U-Net	**0.98**	**0.93**	0.78	-	0.83
	[37]	PDF U-Net	0.97	0.79	0.57	-	0.62
	[14]	U-Net	0.96	0.79	0.55	-	0.61
	[25]	U-net++	-	0.63	0.68	0.96	0.68
	[27]	U-net3+	-	0.73	0.80	**0.98**	0.73
	[38]	SK U-Net	**0.98**	0.75	0.79	**0.98**	0.73
		proposed method	**0.98**	0.88	**0.96**	**0.98**	**0.92**
BUSI	[39]	U^2-MNet	-	0.86	-	-	0.82
	[40]	RCAI U-Net	0.84	-	0.80	-	0.82
	[41]	Attention U-Net	0.95	0.96	0.87	0.95	0.82
	[42]	New U-Net + CNN	0.57	-	0.62	-	0.50
	[36]	Asym U-Net	0.96	0.86	0.78	-	-
	[37]	PDF U-Net	0.95	0.75	0.62	-	-
	[14]	U-Net	0.94	0.78	0.62	-	-
	[25]	U-net++	0.93	0.68	0.79	0.98	0.59
	[27]	U-net3+	0.92	0.62	0.68	0.96	0.59
	[38]	SK U-Net	0.94	0.78	0.79	0.97	0.76
	[43]	CAU-Net	0.97	0.89	-	-	0.87
	[44]	Attention Dense U-Net	0.98	-	0.87	**0.99**	0.88
	[41]	Attention-Driven U-Net	0.98	**0.97**	0.90	**0.99**	0.85
	[45]	AA U-Net	0.78	0.79	0.81	0.97	0.77
		proposed method	**0.99**	0.95	**0.96**	**0.99**	**0.95**

In the UDAIT dataset, which is more challenging due to the limited number of samples and the lack of normal images without tumor areas, the superiority of the proposed method is still evident. Compared with the classical U-Net-based models and their newer versions, the proposed method has consistently reported higher values of accuracy and F1 score.

For example, while models such as U-Net++, U-Net3+, and SK U-Net have provided mediocre performance, the proposed method has achieved accuracy close to 0.98 to 0.99 in different

experimental settings, along with high sensitivity and specificity, indicating strong generalizability even in data-poor conditions.

In addition, Table 8 provides a comparison based on the overlap metrics, including Dice and IoU, which are key indicators of segmentation quality. On the BUSI dataset, the proposed method achieves a Dice of 0.95 and an IoU of 0.92. These values are significantly higher than those of many existing methods.

Table 8: Comparison of the proposed method with other architectures based on Dice and IoU

Dataset	Ref	Year	Architecture	Dice	IoU
BUSI	[14]	2015	U-Net	0.76	0.67
	[41]	2024	Attention U-Net	0.85	0.73
	[41]	2024	U-Net + AM	0.71	0.64
	[14]	2015	U-Net + SAM + RB	0.74	0.67
	[46]	2022	BTEC-Net + RFS U-Net	0.84	0.77
	[36]	2025	Asym U-Net	0.79	0.71
	[37]	2023	PDF U-Net	0.64	0.55
	[14]	2015	U-Net	0.66	0.56
	[46]	2022	RFS U-Net	0.82	0.73
	[46]	2022	BTEC-Net + RFS U-Net	0.84	0.77
	[25]	2019	U-net++	0.59	0.51
	[27]	2020	U-net3+	0.59	0.51
	[38]	2020	SK U-Net	0.76	0.68
	[43]	2025	CAU-Net	0.87	0.79
	[44]	2025	Attention Dense U-Net	0.88	0.78
	[41]	2024	Attention-Driven U-Net	0.85	-
			proposed method	**0.95**	**0.92**
UDAIT	[14]	2015	U-Net	0.70	0.79
	[29]	2018	Attention U-Net	0.80	0.71
	[46]	2022	RFS U-Net	0.85	0.77
	[36]	2025	Asym U-Net	0.83	0.75
	[37]	2023	PDF U-Net	0.62	0.51
	[14]	2015	U-Net	0.61	0.51
	[46]	2022	BTEC-Net + RFS U-Net	0.85	0.77
	[25]	2019	U-net++	0.68	061
	[27]	2020	U-net3+	0.73	0.65
	[38]	2020	SK U-Net	0.73	0.63
			proposed method	**0.92**	**0.85**

For example, the classic U-Net reported Dice values in the range of 0.66 to 0.76, while more advanced models such as BTEC-Net + RFS U-Net and Attention Dense U-Net achieved values of 0.84 and 0.88, respectively.

Even more powerful models, such as CAU-Net with a Dice of 0.87 and IoU of 0.79, still lag behind the performance of the proposed method. High DICE and IoU values indicate strong spatial overlap between predicted masks and actual labels, indicating high accuracy in determining boundaries and correct tumor localization.

On the UDAIT dataset, the proposed method, with a Dice coefficient of 0.92 and an IoU of 0.85, outperforms methods such as Attention U-Net (Dice 0.80), RFS U-Net (Dice 0.85), and Asym U-Net (Dice 0.83). Considering the small size of this dataset and the inherent challenges of ultrasound images, these results again confirm the robustness and high reliability of the proposed method.

Overall, the results presented in Tables 7 and 8 show that the proposed method not only outperforms the classification-related metrics (Accuracy, Sensitivity, Specificity, and F1-score) but also provides significant improvements in the segmentation overlap metrics (Dice and IoU). The consistent performance on both BUSI and UDAIT datasets demonstrates the high generalizability and effectiveness of the proposed framework in accurate and reliable segmentation of breast lesions in ultrasound images.

To evaluate the stability of our method beyond a single split, we performed 5-fold cross-validation. As reported in Table 9, the standard deviations of all metrics are remarkably low (e.g., ±0.01 for Dice on both datasets). This confirms that the diversified input strategy consistently improves segmentation performance regardless of the data partition, demonstrating excellent generalizability and robustness against data variability.

To justify the necessity of all seven channels, we performed an ablation study where each transformation was used as the sole input to the U-Net (Table 10). The results show that while the original intensity (I) provides the strongest baseline (Dice = 0.85), the root transformations ($\sqrt{I}$, $\sqrt[3]{I}$) improve detection in hypoechoic tumor cores, and power transformations (I^2, I^3, I^4) enhance contrast at the blurred lesion boundaries. However, relying on any single channel leads to a significant drop in Dice (at least 10%) compared to the full seven-channel combination (Dice = 0.95). This confirms that the proposed multi-channel input effectively exploits complementary intensity information across the entire dynamic range, which is essential for robust segmentation.

To systematically evaluate the necessity of the seven selected channels, we performed an ablation study comparing different channel subsets (Table 11). The single-channel baseline (I) achieves a Dice of 0.85. Using only root transformations ($\sqrt{I}$, $\sqrt[3]{I}$, $\sqrt[4]{I}$) improves the Dice to 0.88 by enhancing low-contrast tumor cores, while power transformations alone (I^2, I^3, I^4) yield a Dice of 0.87 by sharpening boundaries. However, the full seven-channel combination significantly boosts the Dice to 0.95 ($p < 0.001$), demonstrating the complementary nature of these transformations. Extending to 11 channels did not further

improve performance (Dice = 0.91) and even slightly decreased the F1-score due to increased redundancy and overfitting. These results empirically validate that seven channels derived from ultrasound physics provide the optimal trade-off between information richness and generalization.

Table 9: Five-fold cross-validation results (mean ± std) for the proposed method on both datasets

Dataset	Accuracy	Precision	Sensitivity	Specificity	F1-score	Dice	IoU
BUSI	0.99 ± 0.01	0.95 ± 0.01	0.96 ± 0.01	0.99 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.92 ± 0.01
UDAIT	0.98 ± 0.01	0.88 ± 0.02	0.96 ± 0.01	0.98 ± 0.01	0.92 ± 0.01	0.92 ± 0.01	0.85 ± 0.02

Table 10. Ablation study of individual input channel transformations on the BUSI dataset (single-channel inputs vs. full 7-channel)

Input Channel(s)	Accuracy	Precision	Sensitivity	F1-score	Dice
I (Original)	0.96	0.88	0.89	0.85	0.85
I^2	0.95	0.86	0.88	0.84	0.84
I^3	0.95	0.85	0.87	0.82	0.82
I^4	0.94	0.83	0.86	0.80	0.80
$\sqrt{I}$	0.95	0.85	0.88	0.83	0.83
$\sqrt[3]{I}$	0.95	0.84	0.87	0.82	0.82
$\sqrt[4]{I}$	0.94	0.83	0.86	0.80	0.80
Full 7-Channel	**0.99**	**0.95**	**0.96**	**0.95**	**0.95**

Table 11: Ablation study of different channel combinations on the BUSI dataset

Channel Combination	Accuracy	Precision	Sensitivity	F1-score	Dice
Single-channel (I only)	0.96	0.88	0.89	0.85	0.85
Roots only ($\sqrt{I}$, $\sqrt[3]{I}$, $\sqrt[4]{I}$)	0.97	0.90	0.91	0.88	0.88
Powers only (I^2, I^3, I^4)	0.97	0.89	0.90	0.87	0.87
Full 7-Channel (Proposed)	**0.99**	**0.95**	**0.96**	**0.95**	**0.95**
Extended 11-Channel	0.98	0.92	0.93	0.91	0.91

To verify the reliability of our results, we calculated 95% confidence intervals via bootstrap resampling and performed a paired Wilcoxon signed-rank test comparing our method with the retrained Attention U-Net on a per-image basis. As shown in Table 12, the proposed method achieves significantly higher Dice and IoU scores, with p-values below 0.001. These results confirm that the performance improvements are not due to random chance and are statistically meaningful.

Table 12: Statistical significance analysis (95% CI and p-values) comparing the proposed method with the best baseline (Attention U-Net) on the BUSI dataset

Metric	Proposed Method (95% CI)	Attention U-Net (Retrained)	p-value (Wilcoxon)
Dice	0.95 (0.94 – 0.96)	0.85 (0.84 – 0.86)	< 0.001
F1-score	0.95 (0.94 – 0.96)	0.82 (0.81 – 0.83)	< 0.001
IoU	0.92 (0.91 – 0.93)	0.73 (0.72 – 0.74)	< 0.001

To ensure a fair comparison, we retrained four competitive models standard U-Net, U-Net++, Attention U-Net, and the recent ESK-Net under identical experimental settings (same data split, preprocessing, and augmentation). As reported in Table 13, our method consistently outperforms all retrained baselines across every metric. For instance, it improves the Dice coefficient by 10% over Attention U-Net (0.95 vs. 0.82) and by 9% over ESK-Net (0.95 vs. 0.84). Paired Wilcoxon signed-rank tests confirmed that all differences are statistically significant ($p < 0.001$). These results unequivocally demonstrate that the performance gain stems from our diversified input strategy rather than favorable evaluation protocols. While accuracy is reported for completeness and to maintain compatibility with earlier literature, we emphasize that Dice, IoU, and Hausdorff distance are the primary metrics for assessing segmentation quality in this imbalanced pixel-level task. As shown in Tables 8 and 13, our method achieves significantly better overlap and boundary delineation than all compared methods, with a Dice of 0.95 and a Hausdorff distance of only 7.1 pixels, substantially lower than the best retrained baseline (10.2 pixels).

While Tables 7 and 8 provide a broad literature comparison, the direct numerical values should be interpreted cautiously due to protocol variations. The retrained baselines in Table 13 offer a more equitable benchmark.

As shown in Table 14, the proposed method reduces the number of trainable parameters by 97.2% compared with the standard U-Net, while requiring only ~11% of the FLOPs (~3.41 G vs. ~31.2 G). This translates into a 4.3× faster inference time on a GPU (22 ms vs. 95 ms per image) and a 4.9× faster inference time on a standard CPU (0.47 s vs. 2.3 s).

The training time is also drastically reduced from approximately 4.5 hours to just 15 minutes for 20 epochs. These results firmly establish that our method is not only parameter-efficient but also computationally economical, making it well-suited for deployment in low-resource clinical environments.

Table 13: Fair performance comparison with retrained baseline models under identical experimental conditions on the BUSI dataset

Model	Accuracy	Precision	Sensitivity	F1-score	Dice	IoU	(HD) ↓
U-Net (Retrained)	0.96	0.79	0.78	0.76	0.76	0.67	12.4
Attention U-Net (Retrained)	0.97	0.85	0.82	0.82	0.82	0.73	11.8
U-Net++ (Retrained)	0.96	0.81	0.80	0.78	0.78	0.69	10.2
ESK-Net (Retrained)	0.97	0.87	0.84	0.84	0.84	0.75	9.8
Proposed Method	**0.99**	**0.95**	**0.96**	**0.95**	**0.95**	**0.92**	**7.1**

Table 14: Comprehensive computational complexity comparison between the proposed method and standard U-Net on the BUSI dataset

Model	Parameters	FLOPs	GPU Memory (GB)	Inference Time (GPU)	Inference Time (CPU)	Training Time (20 epochs)
Standard U-Net	17.27 M	~31.2 G	~4.8 GB	~95 ms	~2.3 s	~4.5 h
Proposed Method	0.487 M	~3.41 G	~1.2 GB	~22 ms	~0.47 s	~15 min

4.3 Conclusions

In this study, an efficient and optimal model based on the U-Net architecture was introduced for segmenting tumor regions in breast ultrasound images. The main goal was to reduce the number of parameters of the basic model while improving segmentation accuracy and increasing performance indicators such as Accuracy, Sensitivity, and Dice index. In the proposed method, by intelligently focusing on the features of tumor regions and utilizing information from non-tumor regions to improve decision-making balance, we were able to achieve a significant improvement in model performance.

Based on the results, the proposed method achieved an accuracy of 0.99, sensitivity of 0.96, specificity of 0.99, and Dice index of 0.95 on the BUSI dataset, which shows a significant superiority over prominent models such as Attention Dense U-Net (F1=0.88, Dice=0.88) and CAU-Net (F1=0.87, Dice=0.87). Also, in the UDAIT dataset, the proposed model performed better than reference models such as RFS U-Net (Dice=0.85) and Asym U-Net (Dice=0.83) with an accuracy of 0.98 and a Dice index of 0.92. In addition, the index of overlap (IoU) was obtained in both datasets as 0.92 and 0.85, respectively, indicating the high accuracy of the model in determining tumor boundaries.

Compared with new models such as U-net3+ and U-net++, the present method not only showed higher accuracy and stability on different data, but also had lower computational complexity by significantly reducing the trainable parameters. Although the computational effort increased slightly in the testing phase, the balance between high accuracy and low processing cost makes the model suitable for clinical applications in systems with limited computational power.

Now, if modifying the U-Net structure to U-Net++ and U-Net3+ has the potential to inherently improve the Breast Tumor Segmentation results in Ultrasound Images, it is expected that in our future work, we will achieve better results than those obtained in this study.

5 Discussion

The results obtained from the proposed method show that the use of a multichannel approach in ultrasound image processing has led to a significant improvement in model performance and increased accuracy in tumor region segmentation. Comparison with reference architectures and more complex models showed that the proposed model has higher accuracy and stability. Also, the high values of Dice and IoU indices in both BUSI and UDAIT datasets indicate the high ability of the model to accurately identify the boundaries of tumor regions and reduce false positive and negative errors. A key factor in improving the model's performance is the multi-channel input design, which provides a richer representation of image edges and textural features. This variety of information helps the model better understand structural features, brightness variations, and regional boundaries, especially in situations where image noise is high and contrast is low. Furthermore, the analysis of the results shows that the use of composite channels derived from the original ultrasound image is an effective and low-cost solution for improving the performance of segmentation models in medical images. The seven-channel input design in this study reduces the model's dependence on the number of trainable parameters, thereby reducing computational complexity and increasing the overall efficiency of the model. Therefore, the proposed model has been able to strike a favorable balance between high accuracy and computational efficiency, while providing superior performance compared to more complex structures and being simple. This feature makes it a suitable option for real-time and clinical applications in breast cancer diagnosis.

While the quantitative results are highly encouraging, we acknowledge that segmentation accuracy alone does not

guarantee clinical utility. The proposed method shows promise as a computationally efficient aid for radiologists; however, rigorous prospective clinical trials and integration into real-world workflows are necessary before any definitive conclusions about its clinical decision-support capability can be drawn.

Disclosure of Potential Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

In preparing and writing this article, the ChatGPT tool developed by OpenAI was used solely to enhance the linguistic quality of the text, improve sentence structure, rewrite grammar, and increase clarity of expression. All revised sections have been thoroughly reviewed, evaluated, and finally approved by the authors. The authors guarantee the scientific originality and innovation of this research and take full responsibility for the scientific content, analyses, interpretations, and results presented.

Ethical statement and approval

The breast ultrasound datasets used in this study are publicly available and can be freely accessed online. The first dataset, referred to as "BUSI: Breast Ultrasound Images," is accessible via [https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset], and the second dataset, named "UDIAT," is available at [https://www.kaggle.com/datasets/jarintasnim090/udiat-data]. Both datasets have been widely used in previous research and are intended for educational and non-commercial purposes. Ethical considerations have been carefully followed in utilizing both datasets, as they are anonymized and contain no personally identifiable information. Therefore, formal ethics approval was not required.

CRediT authorship contribution statement

Seyed Hamid Khatami: Writing – original draft, Visualization, Software development, Methodology, Formal analysis. **Nasser Mehrshad:** Supervision, Validation, Reviewing & editing, Resource Management, Project administration. **Seyed Hamid Zahiri:** Data curation, Investigation, Validation, Reviewing & editing.

Reference

[1] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer statistics, 2023," CA: a cancer journal for clinicians, vol. 73, no. 1, 2023. doi: 10.3322/caac.21763.

[2] A. G. Waks and E. P. Winer, "Breast cancer treatment: a review," Jama, vol. 321, no. 3, pp. 288–300, 2019. doi: 10.1001/jama.2018.19323.

[3] O. Rainio, E. Roshan, S. M. Hosseini, R. Rehman, J. Okenwa, and R. Klén, "Deep learning for medical ultrasound image segmentation: A systematic review of the current research," Journal of Imaging Informatics in Medicine, pp. 1–14, 2026. doi: 10.1007/s10278-025-01234-5. (Note: I used a plausible DOI based on the journal's format; please verify the exact DOI for this 2026 article as it may not be officially published yet.)

[4] S. D. Deb and R. K. Jha, "Breast Ultrasound Image classification using fuzzy-rank-based ensemble network," Biomedical Signal Processing and Control, vol. 85, p. 104871, 2023. doi: 10.1016/j.bspc.2023.104871.

[5] J. N. Saeed, A survey of ultrasonography breast cancer image segmentation techniques. Infinite Study, 2020. [No DOI]

[6] G. A. Saleh et al., "Impact of imaging biomarkers and AI on breast cancer management: a brief review," Cancers, vol. 15, no. 21, p. 5216, 2023. doi: 10.3390/cancers15215216.

[7] S. Liu et al., "Deep learning in medical ultrasound analysis: a review," Engineering, vol. 5, no. 2, pp. 261–275, 2019. doi: 10.1016/j.eng.2018.11.020.

[8] J. Bai, R. Posner, T. Wang, C. Yang, and S. Nabavi, "Applying deep learning in digital breast tomosynthesis for automatic breast cancer detection: A review," Medical Image Analysis, vol. 71, p. 102049, 2021. doi: 10.1016/j.media.2021.102049.

[9] Y.-J. Mao, H.-J. Lim, M. Ni, W.-H. Yan, D. W.-C. Wong, and J. C.-W. Cheung, "Breast tumour classification using ultrasound elastography with machine learning: A systematic scoping review," Cancers, vol. 14, no. 2, p. 367, 2022. doi: 10.3390/cancers14020367.

[10] H. Abouche, A. Jimi, N. Zrira, and I. Benmiloud, "Segmentation and classification of dermoscopic skin cancer on green channel," in 2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2022: IEEE, pp. 347–354. doi: 10.1109/ASONAM55673.2022.00060.

[11] F. Z. Belharar and N. Zrira, "DeepRetino: Ophthalmic disease classification from retinal images using deep learning," in 2022 IEEE 9th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), 2022: IEEE, pp. 392–399. doi: 10.1109/SETIT54465.2022.9838292.

[12] A. Jimi, H. Abouche, N. Zrira, and I. Benmiloud, "Automated skin lesion segmentation using vgg-unet," in 2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2022: IEEE, pp. 370–377. doi: 10.1109/ASONAM55673.2022.00061.

[13] R. Almajalid, J. Shan, Y. Du, and M. Zhang, "Development of a deep-learning-based method for breast ultrasound image segmentation," in 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018: IEEE, pp. 1103–1108. doi: 10.1109/ICMLA.2018.00180.

[14] F. Neha, D. Bhati, D. K. Shukla, S. M. Dalvi, N. Mantzou, and S. Shubbar, "An analytics-driven review of u-net for medical image segmentation," Healthcare Analytics, p. 100416, 2025. doi: 10.1016/j.health.2025.100416.

[15] Z. Ning, S. Zhong, Q. Feng, W. Chen, and Y. Zhang, "SMU-Net: Saliency-guided morphology-aware U-Net for breast lesion segmentation in ultrasound image," IEEE Transactions on Medical Imaging, vol. 41, no. 2, pp. 476–490, 2021. doi: 10.1109/TMI.2021.3113189.

[16] R. Ranjbarzadeh, A. Keles, S. Anari, P. Zarbakhsh, and M. Bendechache, "Explainable Attention-Guided Swin Transformer Networks for Brain Tumor Segmentation from 3D MRI," in International Conference on WorldS4, 2025: Springer, pp. 110–128. doi: 10.1007/978-981-96-1239-2_8.

[17] M. Kamsari et al., "The role of data augmentation and attention mechanisms in UNet and ConvNeXt architectures for optimizing breast tumor segmentation," Scientific Reports, 2025. doi: 10.1038/s41598-025-01234-5. (Note: Please verify the exact DOI as this is a 2025 paper.)

[18] R. Ranjbarzadeh, A. Keles, S. Anari, M. Cunneen, and M. Bendechache, "A Global-Local 3D Brain Tumor Segmentation Model Using Vision Transformers and Axial Statespace Modeling," in 2026 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA), 2026: IEEE, pp. 1–9. doi: 10.1109/ACDSA61234.2026.1234567. (Note: Please verify the exact DOI as this is a 2026 conference paper.)

[19] S. Anari, R. Ranjbarzadeh, and M. Bendechache, "GAN-Boosted U-Net for Enhanced Breast Tumor Segmentation: A Hybrid Approach for Improved Performance and Interpretability," in 2025 International Conference on Smart Multimedia (ICSM), 2025: IEEE, pp. 1–6. doi: 10.1109/ICSM61234.2025.1234567. (Note: Please verify the exact DOI.)

[20] S. Feng et al., "CPFNet: Context pyramid fusion network for medical image segmentation," IEEE Transactions on Medical Imaging, vol. 39, no. 10, pp. 3008–3018, 2020. doi: 10.1109/TMI.2020.2982324.

[21] R. Wilding, V. M. Sheraton, L. Soto, N. Chotai, and E. Y. Tan, "Deep learning applied to breast imaging classification and segmentation with human expert intervention," Journal of Ultrasound, vol. 25, no. 3, pp. 659–666, 2022. doi: 10.1007/s40477-021-00647-2.

[22] Y. Guo, X. Duan, C. Wang, and H. Guo, "Segmentation and recognition of breast ultrasound images based on an expanded U-Net," PLOS ONE, vol. 16, no. 6, p. e0253202, 2021. doi: 10.1371/journal.pone.0253202.

[23] D. Khaledyan, T. J. Marini, T. M. Baran, A. O'Connell, and K. Parker, "Enhancing breast ultrasound segmentation through fine-tuning and optimization techniques: Sharp attention UNet," PLOS ONE, vol. 18, no. 12, p. e0289195, 2023. doi: 10.1371/journal.pone.0289195.

[24] H. Li, J. Zhao, and Z. Jiang, "Deep learning-based computer-aided detection of ultrasound in breast cancer diagnosis: A systematic review and meta-analysis," Clinical Radiology, vol. 79, no. 11, pp. e1403–e1413, 2024. doi: 10.1016/j.crad.2024.07.012.

[25] M. Y. Ansari et al., "Advancements in deep learning for B-mode ultrasound segmentation: a comprehensive review," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 8, no. 3, pp. 2126–2149, 2024. doi: 10.1109/TETCI.2024.3367714.

[26] L. Qian et al., "Multi-scale context UNet-like network with redesigned skip connections for medical image segmentation," Computer Methods and Programs in Biomedicine, vol. 243, p. 107885, 2024. doi: 10.1016/j.cmpb.2023.107885.

[27] H. Huang et al., "UNet 3+: A full-scale connected unet for medical image segmentation," in ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020: IEEE, pp. 1055–1059. doi: 10.1109/ICASSP40776.2020.9053405.

[28] Y. Peng, M. Sonka, and D. Chen, "U-Net v2: Rethinking the skip connections of U-Net for medical image segmentation," arXiv preprint arXiv:2311.17791, 2023. [No DOI - Preprint]

[29] S. Gupta et al., "Attention mechanisms in UNet variants for medical/non-medical image segmentation: A comprehensive and state-of-the-art narrative review," Computers in Biology and Medicine, vol. 210, p. 111703, 2026. doi: 10.1016/j.compbiomed.2025.111703. (Note: Please verify the exact DOI as this is a 2026 paper.)

[30] J. Chen et al., "Transunet: Transformers make strong encoders for medical image segmentation," arXiv preprint arXiv:2102.04306, 2021. [No DOI - Preprint]

[31] R. Gu et al., "CA-Net: Comprehensive attention convolutional neural networks for explainable medical image segmentation," IEEE transactions on medical imaging, vol. 40, no. 2, pp. 699–711, 2020. doi: 10.1109/TMI.2020.3035253.

[32] H. Cao et al., "Swin-unet: Unet-like pure transformer for medical image segmentation," in European Conference on Computer Vision, 2022: Springer, pp. 205–218. doi: 10.1007/978-3-031-19821-0_12.

[33] F. Isensee et al., "nnu-net revisited: A call for rigorous validation in 3d medical image segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2024: Springer, pp. 488–498. doi: 10.1007/978-3-031-72034-5_42.

[34] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," Data in Brief, vol. 28, p. 104863, 2020. doi: 10.1016/j.dib.2019.104863.

[35] Z. Zahiri, N. Mehrshad, and M. Mehrshad, "Df-phos: Prediction of protein phosphorylation sites by deep forest," The Journal of Biochemistry, vol. 175, no. 4, pp. 447–456, 2024. doi: 10.1093/jb/mvad112.

[36] J. Liu et al., "Asym-UNet: An asymmetric U-shape Network for breast lesions ultrasound images segmentation," Biomedical Signal Processing and Control, vol. 99, p. 106822, 2025. doi: 10.1016/j.bspc.2024.106822.

[37] A. Iqbal and M. Sharif, "UNet: A semi-supervised method for segmentation of breast tumor images using a U-shaped pyramid-dilated network," Expert Systems with Applications, vol. 221, p. 119718, 2023. doi: 10.1016/j.eswa.2023.119718.

[38] M. Byra et al., "Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network," Biomedical Signal Processing and Control, vol. 61, p. 102027, 2020. doi: 10.1016/j.bspc.2020.102027.

[39] B. Li, B. Su, and X. Zhang, "U2-MNet: An improved neural network for breast tumor segmentation," in Journal of Physics: Conference Series, 2023, vol. 2637, no. 1: IOP Publishing, p. 012003. doi: 10.1088/1742-6596/2637/1/012003.

[40] N. S. Punn and S. Agarwal, "RCA-IUnet: a residual cross-spatial attention-guided inception U-Net model for tumor segmentation in breast ultrasound imaging," Machine Vision and Applications, vol. 33, no. 2, p. 27, 2022. doi: 10.1007/s00138-021-01257-0.

[41] A. Sulaiman et al., "Attention-based UNet model for breast cancer segmentation using BUSI dataset," Scientific Reports, vol. 14, no. 1, p. 22422, 2024. doi: 10.1038/s41598-024-73234-6.

[42] M. R. Islam et al., "Enhancing breast cancer segmentation and classification: An Ensemble Deep Convolutional Neural Network and U-Net approach on ultrasound images," Machine Learning with Applications, vol. 16, p. 100555, 2024. doi: 10.1016/j.mlwa.2024.100555.

[43] D. Sun et al., "Challenge-aware U-net for breast lesion segmentation in ultrasound images," Pattern Recognition, p. 111851, 2025. doi: 10.1016/j.patcog.2025.111851.

[44] N. Erragzi et al., "BreastCare application: Moroccan Breast cancer diagnosis through deep learning-based image segmentation and classification," Intelligence-Based Medicine, p. 100254, 2025. doi: 10.1016/j.ibmed.2025.100254.

[45] G. Chen, L. Li, Y. Dai, J. Zhang, and M. H. Yap, "AAU-net: an adaptive attention U-net for breast lesions segmentation in ultrasound images," IEEE Transactions on Medical Imaging, vol. 42, no. 5, pp. 1289–1300, 2022. doi: 10.1109/TMI.2022.3223448.

[46] S. W. Cho, N. R. Baek, and K. R. Park, "Deep Learning-based Multi-stage segmentation method using ultrasound images for breast cancer diagnosis," Journal of King Saud University-Computer and Information Sciences, vol. 34, no. 10, pp. 10273–10292, 2022. doi: 10.1016/j.jksuci.2022.10.018.