RESEARCH ARTICLE

# Enhancing breast ultrasound segmentation through fine-tuning and optimization techniques: Sharp attention UNet

**Donya Khaledyan**[1]*, **Thomas J. Marini**[2], **Timothy M. Baran**[2], **Avice O'Connell**[2], **Kevin Parker**[1,2]

**1** Department of Electrical and Electronics Engineering, University of Rochester, Rochester, NY, United States of America, **2** Department of Imaging Sciences, University of Rochester Medical Center, Rochester, NY, United States of America

* dkhaledy@ur.rochester.edu
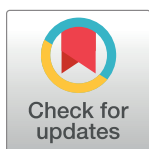
## Abstract

Segmentation of breast ultrasound images is a crucial and challenging task in computer-aided diagnosis systems. Accurately segmenting masses in benign and malignant cases and identifying regions with no mass is a primary objective in breast ultrasound image segmentation. Deep learning (DL) has emerged as a powerful tool in medical image segmentation, revolutionizing how medical professionals analyze and interpret complex imaging data. The UNet architecture is a highly regarded and widely used DL model in medical image segmentation. Its distinctive architectural design and exceptional performance have made it popular among researchers. With the increase in data and model complexity, optimization and fine-tuning models play a vital and more challenging role than before. This paper presents a comparative study evaluating the effect of image preprocessing and different optimization techniques and the importance of fine-tuning different UNet segmentation models for breast ultrasound images. Optimization and fine-tuning techniques have been applied to enhance the performance of UNet, Sharp UNet, and Attention UNet. Building upon this progress, we designed a novel approach by combining Sharp UNet and Attention UNet, known as Sharp Attention UNet. Our analysis yielded the following quantitative evaluation metrics for the Sharp Attention UNet: the Dice coefficient, specificity, sensitivity, and F1 score values obtained were 0.93, 0.99, 0.94, and 0.94, respectively. In addition, McNemar's statistical test was applied to assess significant differences between the approaches. Across a number of measures, our proposed model outperformed all other models, resulting in improved breast lesion segmentation.

## 1. Introduction

Breast cancer is a major public health concern, ranking second in cancer-related deaths among women globally, and it is estimated that 1 in 8 females in the United States will develop breast cancer in their lifetime [1–4]. Early detection and treatment are vital for preventing

metastasis and improving survival rates. Screening mammography has been successful in reducing mortality by enabling early detection [5]. However, most of the world currently lacks access to any form of medical imaging for the evaluation of breast masses [6–9]. Ultrasound is a low-cost, and portable imaging modality that provides first-line evaluation for palpable breast lumps. Still, its deployment has been limited by the need for an expert sonographer to acquire the images and a specialist to interpret them. Artificial intelligence (AI) is a promising avenue to circumvent the availability of experienced experts. The value of breast ultrasound segmentation of lesions plays a potentially vital role in diagnosing abnormalities [10]. Moreover, accurately outlining and identifying lesions in ultrasound images could allow the development of practical AI solutions which could aid in early detection and treatment [11, 12]. Additionally, AI can be a replacement solution in instances where there is a lack of specialists; it can also aid specialists in better diagnosis.

Over the past two decades, extensive research has been conducted to develop effective approaches for achieving precise segmentation [13–16]. Deep learning (DL) models, predominantly convolutional neural networks (CNNs), have proven remarkably successful in accurately segmenting anatomical structures and identifying pathological regions in various medical imaging modalities, including X-ray, MRI, CT, and ultrasound [17–26].

Jabeen et al. [27] presented a novel framework for breast cancer classification from mammography images. They fine-tuned the EfficientNet-b0 network [28]. The innovation of this work lies in the combination of contrast enhancement, feature fusion, and a novel feature selection technique, resulting in improved accuracy. However, the increase in computational time due to the feature fusion step may be a consideration for practical implementation. Chaudhury et al. [29] presented a novel approach to texture classification in invasive ductal carcinoma (IDC) using transfer learning with super convergence. The study introduces a lightweight model named Squeeze Net [30] for easy deployment on mobile devices and incorporates data augmentation and color normalization techniques to improve performance. A Grad-CAM based solution is utilized for interpretable feature extraction. The proposed method achieves state-of-the-art results in IDC texture classification and offers potential applications for tumor growth analysis.

Ultrasound's safety, portability, cost-effectiveness, and non-invasiveness, coupled with robust machine learning (ML) algorithms, can potentially enhance early detection, assessment accuracy, and treatment outcomes for breast cancer patients. However, ultrasound images can include high variations in terms of image quality and artifacts [31]. Optimizing and fine-tuning ML models enables them to adapt and perform well across different variations and enhances applicability in real-world scenarios. Moreover, with the increase in data and model complexity, optimization and fine-tuning play a vital and more challenging role than before [32–36].

The fundamental challenge of an ML model is to configure the model or determine the algorithm used for minimizing the loss function. The choice of hyper-parameter values directly impacts the performance of the ML model [32].To achieve an optimal ML model, it is essential to explore a range of possibilities. This process, known as hyperparameter tuning [37], involves designing the ideal model architecture while finding the optimal configuration for the hyperparameters [32, 38].

In recent years, UNet-based models have shown significant advancements in medical image segmentation. The UNet architecture, with its encoder-decoder design and skip connections, has proven to be highly effective in accurately segmenting anatomical structures and lesions in medical images. The UNet architecture, proposed by Ronneberger et al. [39], resolves the issue of decreased fine-grained spatial details in the decoder section by incorporating skip connections that directly link corresponding layers of the encoder and decoder [40].

Over the years, researchers have developed several networks based on UNet architecture, each with unique enhancements and improvements. Some of these networks include Sharp UNet [41], Attention UNet [42], UNet++ [43], UNet3+ [44], CSM-Net [45], Asymmetric UNet [46], Kernel UNet [47], and Swin_UNet [26]. The CSM-Net, Kernel UNet, and Asymmetric UNet are models designed for ultrasound segmentation, with their architecture specifically emphasizing the integration of attention mechanisms.

In this study, we explore the impact of different optimization and fine-tuning techniques on the segmentation results of UNet based models; these techniques include but are not limited to the activation function, loss function, input size, batch size, weight initialization, learning rate schedules, and early stopping on the overall performance of the segmentation models. These techniques aim to improve convergence speed, alleviate overfitting and underfitting, and enhance the model's generalization capabilities [33]. We also tested the performance of UNet, Sharp UNet, Attention UNet, UNet ++, and a novel that we developed known as Sharp Attention UNet. We hypothesized that Sharp Attention UNet would have the best performance compared to the mentioned networks because the combination of salient features as the outcome of attention gates and the sharpened features will improve the network's ability to extract clinically important features.

The results of this study are expected to provide valuable insights into the optimization process for breast ultrasound image segmentation models. By identifying the most effective combinations of activation functions, loss functions, and other optimization techniques the reliability and efficacy of segmentation models will improve leading to better clinical decision-making and patient care.

## 2. Dataset and data preprocessing

### 2.1 Data collection

The dataset we used in this study is known as "Breast Ultrasound Images" (BUSI) [48]. The dataset consists of 780 images from 600 female patients between the ages of 25 and 75, collected in 2018. They were scanned using a LOGIQ E9 ultrasound system and lesions were segmented with manually outlined masks from the radiologist's evaluation. The images are classified into three groups: (1) 133 normal images without masses, (2) 437 images with benign masses, and (3) 210 images with malignant masses. The images are in PNG format, have varying heights and widths, and an average size of 600*500 pixels. The data was preprocessed by removing non-image text and labels.

### 2.2 Image enhancement

The accuracy of DL models can be enhanced by employing image enhancement techniques prior to feeding them to the network [49–52]. Ultrasound images often contain multiple artifacts such as speckle noise, attenuation effect, and low contrast, leading to diminished image quality and interpretation challenges. Enhancing the image quality clarifies patterns, enabling the DL model to identify and classify features within the image more accurately.

In this study, we utilized contrast limited adaptive histogram equalization (CLAHE) [53] as an image enhancement technique. CLAHE is an improvement over adaptive histogram equalization (AHE) [54], and it mitigates the problem of excessive contrast levels in AHE. Unlike AHE, which can exceedingly enhance contrast, CLAHE sets a constraint on the contrast using a histogram. The objective of CLAHE is to enhance image contrast while preserving image quality. This process requires operation on localized image regions, referred to as "tiles" to align the contrast within each tile with a specified histogram shape. To achieve a unified and continuous output image, neighboring tiles are merged using bilinear interpolation. In our

study, we explored various image enhancement techniques, including gamma correction and fuzzy techniques. However, given our dataset and architecture, we ultimately selected CLAHE. We applied it as a preprocessing step and enhancement technique for specific reasons. Firstly, CLAHE excels at enhancing local contrast in images, a valuable trait for tasks like medical image segmentation. By employing CLAHE, we aimed to enhance subtle features and amplify image details, bolstering the segmentation model's ability to capture crucial patterns and boundaries. Furthermore, CLAHE's adaptability is pivotal for medical images, which often exhibit contrast variations across different regions. This adaptability is crucial to address the challenge of uneven illumination and contrast in our dataset, a common occurrence in medical imaging, particularly ultrasound imaging. The CLAHE method can be formulated as Eq 1, where $\beta$ is clip limit, $M$ is area size, $N$ is number of grey-level values, $\alpha$ is clip factor, and $S_{\max}$ is maximum tolerable slope.

$$\beta = \frac{M}{N}\left(1 + \frac{\alpha}{100}\left(S_{\max} - 1\right)\right) \tag{1}$$

### 2.3 Data augmentation

Data augmentation involves modifying existing training data to generate new samples, thereby increasing the training dataset's size. This technique mitigates overfitting by introducing diversity, enhancing the model's performance, and promoting generalization. Particularly in fields like medical imaging, where data scarcity is a concern, data augmentation attempts to create a diverse training set mirroring real-world scenario; this will aid the model's generalization to new, unseen data. However, while augmentation diversifies, generalizes, and balances data, caution must be employed regarding augmentation parameters' impact on diagnostic accuracy. In ultrasound images, extreme brightness or zoom adjustments may lead to the loss of crucial details and distortions that compromise model predictions' accuracy, as ultrasound artifacts, e.g., shadowing, contain vital diagnostic information [55]. Hence, meticulous selection and adjustment of augmentation parameters are imperative to prevent image quality degradation or essential information loss. Moreover, augmentation parameter choices differ based on the image data type. For ultrasound images, techniques preserving underlying structures while introducing variability to counter overfitting are preferred. Subtle rotations or shifts might be more suitable than drastic brightness or zoom changes, maintaining diagnostic relevance [56, 57].

The parameters used in this study include 45-degree rotation, which rotates the image randomly in the range of [–45,45] degrees. The [-0.08,0.08] zoom range randomly zooms in or out of the image. The horizontal flip is on, which flips the randomly selected images horizontally. Moreover, width shift range and height shift range are both 0.15, it shifts the image horizontally and vertically, respectively. The shear range is [-0.03,0.03], it will apply the shear transformation to the image, and the brightness range changes between 0.99 and 1.07, which adjusts the brightness of the image. These parameters are used to define the data generators, which are functions that generate batches of augmented data samples during the training of an ML model. By randomly applying these parameter values to the training data, the resulting augmented dataset better represents real-world scenarios and reduces overfitting of the model to specific instances in the training set.

## 3. Materials and methods

### 3.1 Model optimization techniques

In this section, we focus on exploring various optimization techniques, including activation functions, dropout, and loss functions to fine-tune our segmentation model. By carefully

selecting and fine-tuning these optimization components, we aim to improve the model's ability to accurately outline masses in breast ultrasound images while identifying images with no mass.

**3.1.1 Activation function.** Activation functions are essential parts of neural networks. Nonlinear activation functions allow neural networks to model complex nonlinear relationships between inputs and outputs, which is necessary for many real-world applications. They will ensure that gradients can be propagated through the neural network during backpropagation. The derivative of the activation function regarding its input defines the gradient flow through the network, and different activation functions can affect the stability and speed of gradient propagation.

Rectified Linear Unit (ReLU) is a simple and widely used activation function that returns the input value if it is positive, and zero otherwise [58]. It has been shown to be effective in many neural network architectures due to its simplicity and speed. However, one potential downside of ReLU is that it can suffer from the "dying neurons" problem, where a large portion of the network's neurons can become non-responsive and "die" during training. To address the "dying neurons" problem, the Leaky ReLU (LReLU) activation function was introduced. This function is similar to ReLU but returns a small negative value instead of zero for negative inputs. This ensures that all neurons are active during training, which can lead to better performance. However, with all the advantages over ReLU and the previous activation functions, LReLU still has a sharp edge at 0 value. This can cause optimization issues during training, particularly for gradient-based optimization methods. The sharp edge can lead to abrupt changes in the function's output, which can result in vanishing or exploding gradients. These problems can make the optimization process unstable, slow down training, or even prevent the model from converging altogether.

To overcome this problem, swish was proposed. Swish is a newer activation function that has gained popularity in recent years. It was invented at Google Brain in 2017 by Ramachandran et al. [59]. It is a smooth, non-monotonic function that is similar to the sigmoid function. Swish has been shown to outperform ReLU and LReLU in some neural network architectures, but its performance can be sensitive to hyperparameters. While swish has shown effective performance on certain datasets and architectures, it is not a universal solution that works well for all neural networks and tasks. Swish activation function's equation is:

$$f(x) = x.\text{sigmoid}(\beta x) \tag{2}$$

which $\beta$ is a learnable parameter. However, most implementations exclude the use of this trainable parameter, resulting in the activation function being:

$$f(x) = x.\text{sigmoid}(x) \tag{3}$$

where sigmoid($x$) is:

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \tag{4}$$

Another novel activation function that has demonstrated effectiveness in various applications is mish. Similar to swish, mish is a non-monotonic activation function for deep neural networks, proposed by Diganta Misra in 2019 [60].

The mish activation function is defined as Eq 5:

$$f(x) = x*\tanh\left(\ln(1 + \exp(x))\right) \tag{5}$$

where $x$ is the input to the activation function. The mish function is a smooth, continuous, and

non-monotonic function that is symmetric around the origin. It has a maximum value of 1.0 at $x = 0$ and asymptotes to linearity for very small and very large values of $x$. The mish activation function offers several advantages when compared to other activation functions. One advantage is its ability to enhance the performance of deep neural networks while exhibiting self-regularization. This property mitigates the risk of overfitting by managing the growth of gradients during training. Additionally, the mish activation function is computationally efficient and straightforward to implement, requiring only a small number of elementary operations.

As illustrated in Fig 1(A) mish and swish both possess non-monotonicity, smoothness, and the ability to retain a small quantity of negative weights. These characteristics are responsible for the reliable performance and enhancement of deep neural networks. In the optimization process of DL networks, the first and second derivatives of the activation function can provide critical information about the shape and direction of the function. By analyzing these derivatives, we can determine the direction of steepest descent and whether the function is concave or convex. In the case of mish and swish activation functions, their distinctive negative curvature and smoothness, as evidenced by their first and second derivatives in Fig 1(B), allows for more efficient optimization during the gradient descent process. This can lead to faster convergence and better performance of DL networks.

**3.1.2 Dropout.** Dropout is a regularization method exploited in neural networks to reduce overfitting. Overfitting occurs when a model becomes too complicated compared to the number of data available for training and starts to fit noise in the training data, rather than the underlying patterns [61]. This leads to poor generalization and high error rates on new, unseen data [62].

Dropout avoids overfitting by randomly dropping out a fraction of the neurons in a layer during training. This drives the remaining neurons to learn more robust features that are not dependent on the presence of specific neurons. By doing so, the model becomes less sensitive to the specific details of the training data and is more likely to generalize well to new data. Furthermore, dropout can reduce the effect of co-adaptation between neurons. When neurons co-adapt, they tend to learn analogous features and can become excessively specialized, which can lead to overfitting. There have been many studies regarding the optimal values for dropout. Selecting the optimized drop out values depends on various factors such as the dataset, network architecture, and training method used.

Gal et al. [63] explored the impact of dropout probability and the number of neurons on the performance of deep neural networks. They suggested that higher dropout probabilities are generally better for deeper layers of the network and that the optimal dropout probability may vary based on the specific task and dataset. The optimal value for dropout in this work is 0.1 for the encoder and 0.5 for the decoder. These dropout values were selected based on the recommended optimal values provided in reference [63] and were further fine-tuned through an iterative process of trial-and-error around these values.

**3.1.3 Loss function.** The loss function utilized in this study is a custom loss function for training neural networks for segmentation tasks (Eq 6). It combines two loss functions, the Binary Cross-Entropy (BCE) loss and the dice loss. The BCE loss is a common loss function deployed for binary classification problems, such as image segmentation, where each pixel is classified as either foreground or background. The *BCE* loss measures the difference between the predicted probabilities and the true binary labels (Eq 7).

The Dice loss, on the other hand, measures the overlap between the predicted segmentation mask and the true mask. It is calculated as Eq 8, where the Dice coefficient (Eq 9) measures the similarity between the predicted and true masks [64].
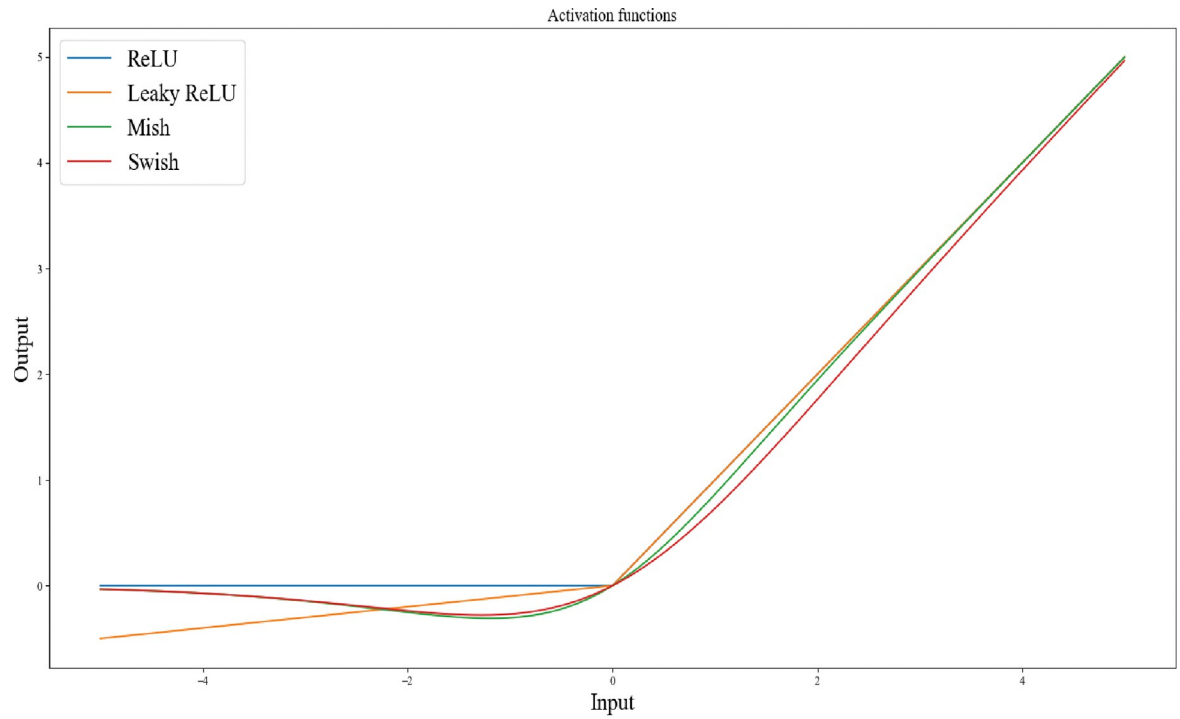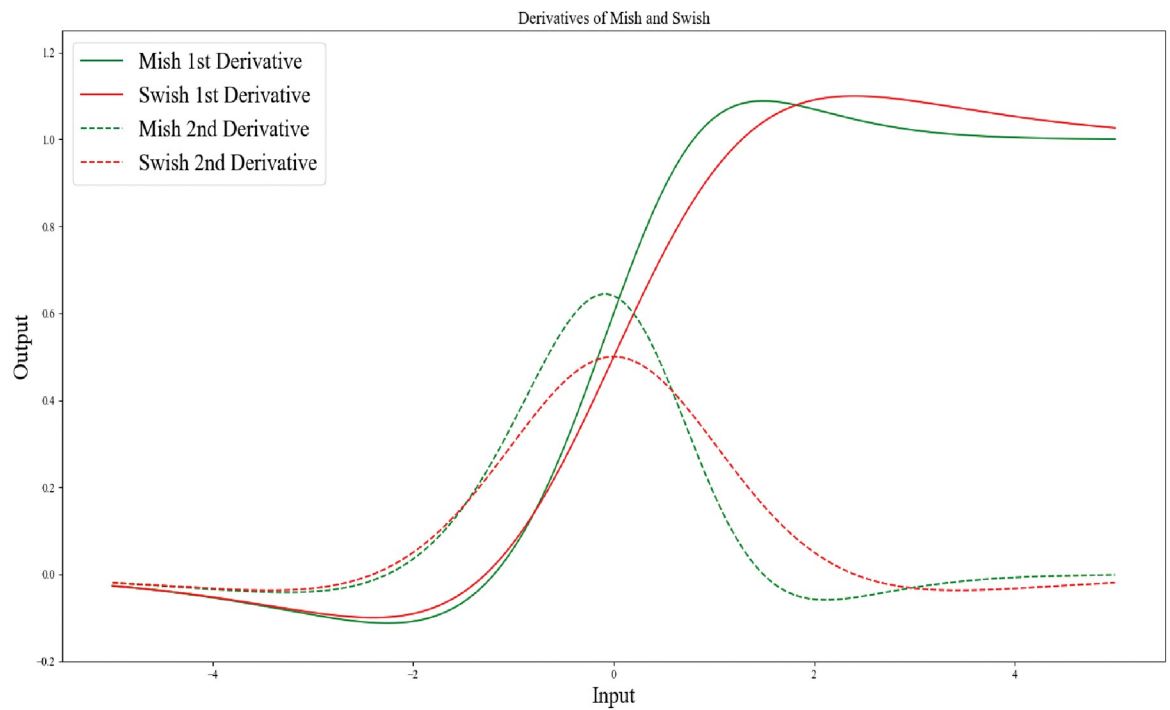
**Fig 1.** (a) Graph of ReLU, Leaky ReLU, mish, and swish. Both mish and swish have a unique negative curvature, setting them apart from ReLU, and Leaky ReLU. b) The 1st and 2nd derivatives of mish and swish activation functions (graphs are plotted in VS Code).

https://doi.org/10.1371/journal.pone.0289195.g001

By combining these two loss functions, the model is encouraged to produce segmentation masks that are both accurate and have high overlap with the true masks. This can lead to better performance on segmentation tasks, particularly when dealing with complex or ambiguous boundaries between foreground and background classes. The equations for $BCE_{\text{Dice}}$ loss is defined as follows:

$$BCE_{\text{Dice}}\text{loss} = BCE \text{ loss} + \text{Dice loss} \tag{6}$$

where $BCE$ loss is calculated as:

$$BCE \text{ loss} = -\frac{1}{N}\sum_{i=0}^{N}[y_i*\log(p_i) + (1 - y_i)*\log(1 - p_i)] \tag{7}$$

where $y$ is the ground truth label (either 0 or 1), $p$ is the predicted probability of the positive class (i.e., the probability of the pixel being part of the object), and log is a natural logarithm function. The Dice loss is calculated as:

$$\text{Dice loss} = 1 - \text{Dice coefficient} \tag{8}$$

$$\text{Dice coefficient} = \frac{(2*\text{intersection})}{(\text{size of prediction} + \text{size of ground truth} + \text{epsilon})} \tag{9}$$

where intersection is the number of pixels where the predicted and ground truth masks both have a value of 1, union is the number of pixels where either the predicted or ground truth mask has a value of 1, and epsilon is a small value (e.g., 1e-5) to avoid division by zero.

From the equations, with $BCE$ loss we can handle training when the foreground and background classes are imbalance, and since the size of the segmented area of the lesions is smaller than background, $BCE$ loss is an appropriate way to tackle that. In the other hand, Dice loss is more sensitive to the foreground which means segmentation accuracy, that would be rational to take advantage of that for an appropriate segmentation. Thus, $BCE_{\text{Dice}}$loss was employed in this study because it combines these two terms to penalize false positives and false negatives (through the $BCE$ loss term) while also encouraging overlap between the predicted and ground truth masks (through the Dice loss term). The final $BCE_{\text{Dice}}$loss is the sum of these two terms.

## 3.2 Models and network architectures

In this section, we explore the models we used with the optimization and fine-tuning steps introduced earlier. Each architecture in Figs 2–4 are designed based on the highest score hyperparameters and network fine tunings.

The proposed model was developed using Python 3.7 with Keres API and was trained and tested on the NVIDIA RTX A3000 graphics processing unit (GPU), and the NVIDIA Tesla P100 GPU. The model underwent training for 300 epochs using the Adam optimizer (learning rate = 0.001, beta1 = 0.9, beta2 = 0.9, epsilon = 0.0000001).

The backbone of the architectures explored in this section are based on UNet. The UNet architecture is a powerful and flexible model for medical image segmentation. Its ability to handle limited training data, accurately identify regions of interest, handle variations and noise, and preserve spatial information makes it highly suitable for a wide range of medical imaging applications [65]. The UNet architecture is based on the fully convolutional network approach, which enables end-to-end pixel-wise predictions. It consists of two main parts: the contracting path (encoder) and the expansive path (decoder). These two parts are connected to each other by skip connections. These skip connections help in preserving the spatial details
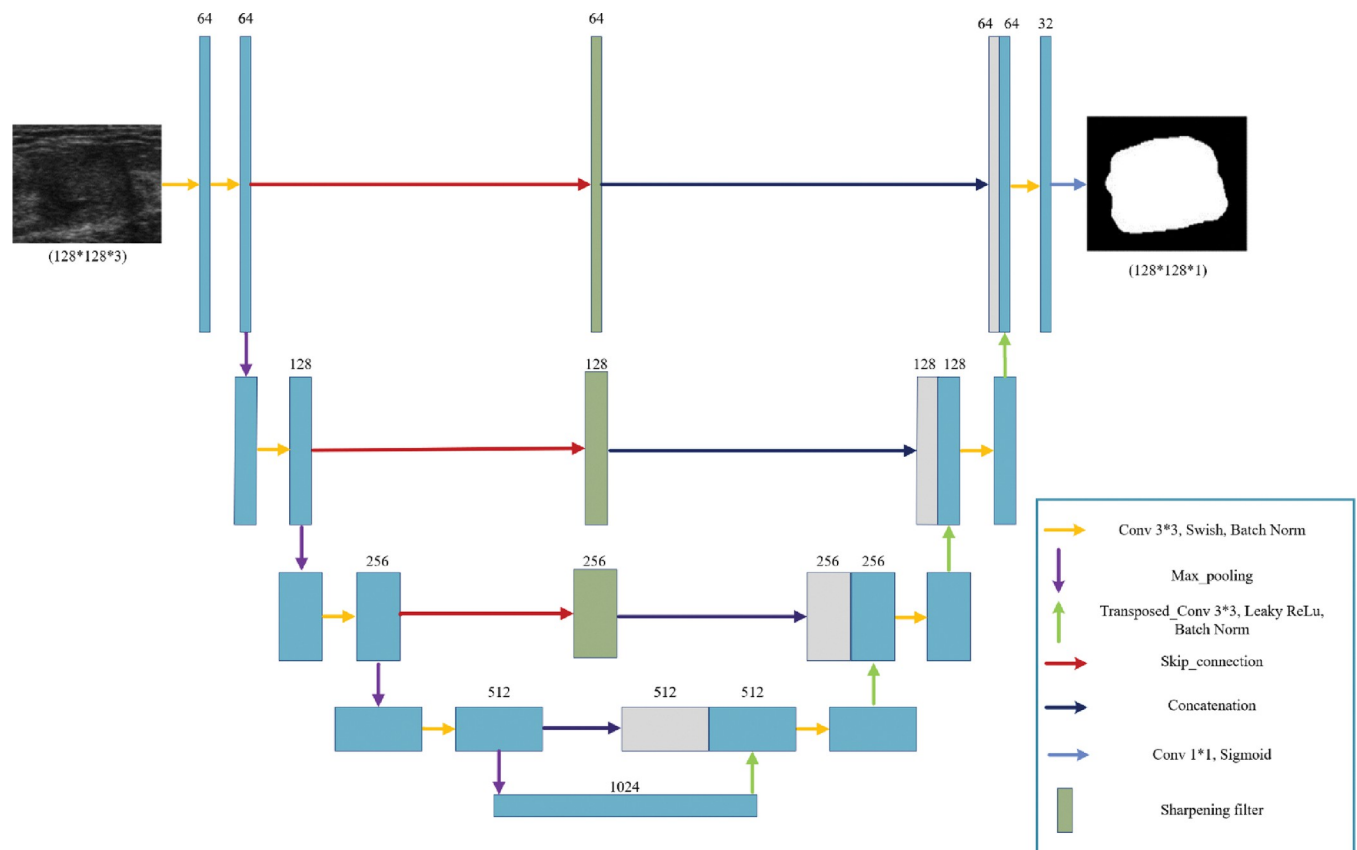
**Fig 2. The Sharp UNet architecture inspired by [41] includes a schematic layout where encoder features are convolved with a sharpening spatial kernel before merging with the decoder features.** This helps to reduce feature mismatches without adding extra parameters or computational cost.

https://doi.org/10.1371/journal.pone.0289195.g002

that are usually lost during the downsampling process in the encoder. By combining the information from the encoder and decoder through concatenation, the spatial information is preserved while enhancing the depth of the feature maps. This enables the network to learn and understand the spatial relationships between different features more effectively.

**3.2.1 Sharp UNet.** The simple skip connections used between the encoder and decoder within a UNet model can cause the gradients to vanish [66], which hampers the model's capacity to accurately segment objects in images. Additionally, these simple skip connections may extract redundant low-level features and fail to capture multi-scale features or selectively focus on important regions in the input image. One approach to handle the loss of information due to the simple skip connections is applying sharpening filter on the encoder path features then concatenate the results with the decoder path [41]. The use of the sharpening filter layer allows for semantically fusing less dissimilar features.

In order to minimize the disparity between encoder and decoder features, in the Sharp UNet a sharpening filter applied within each pathway connecting the encoder and decoder. Image sharpening highlights intensity transitions in an image by using convolution with specific kernels or masks, such as the Laplacian filter. This filter captures changes in intensity across the image and enhances image details.

Fig 2 shows the sharp UNet architecture used in this paper. Among different Laplacian kernels, the kernel given in Eq 10 presented the best performance in our application. Moreover, we observed an enhancement in the performance of our model by eliminating the final sharpening filter that was originally between the last encoder layer and the initial decoder
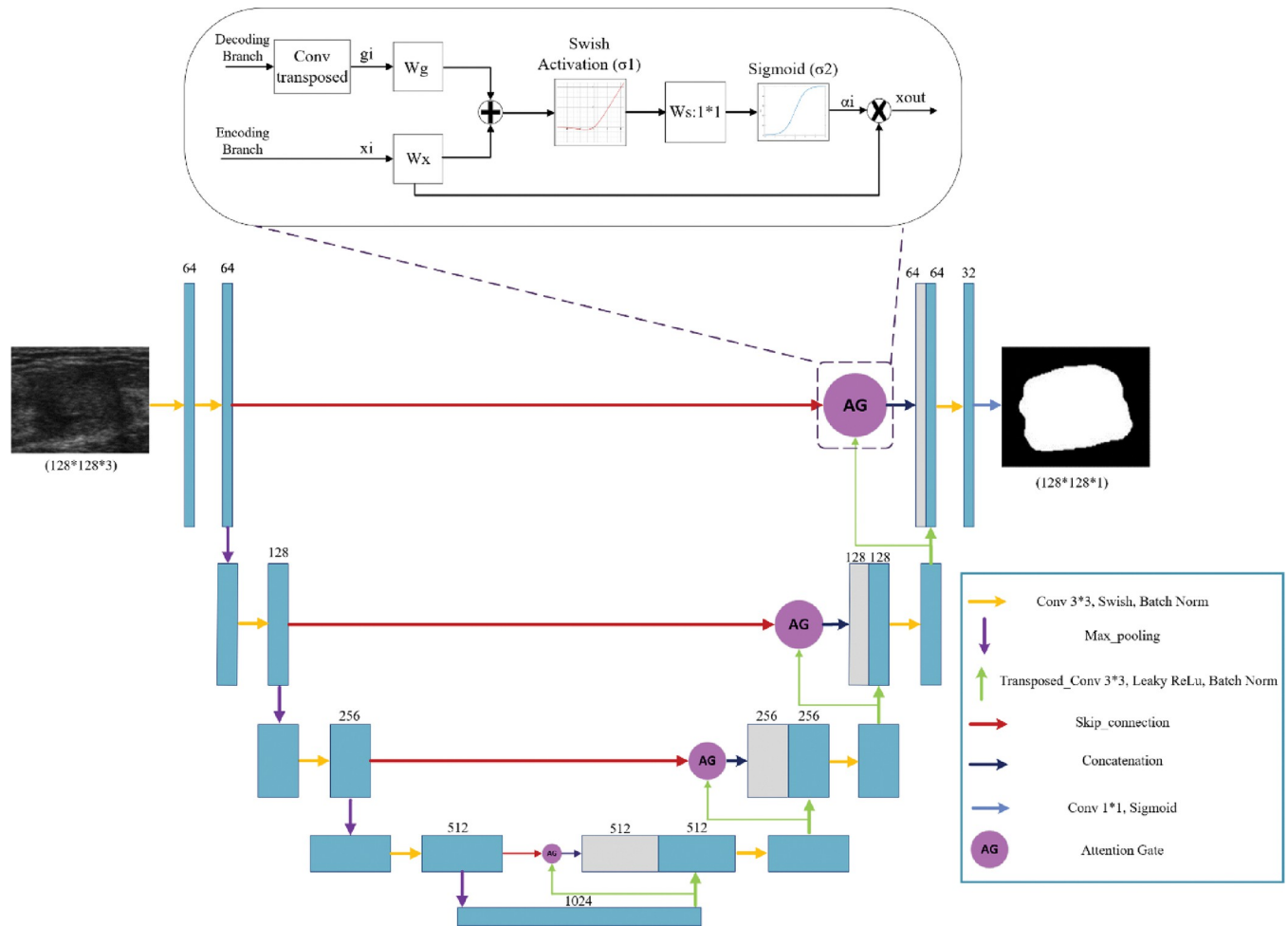
**Fig 3.** Bottom: The attention UNet architecture inspired by [42] the AG selectively emphasizes important features and suppresses irrelevant features. Top: Illustration of the proposed additive Attention Gate (AG), which employs a gating signal (g) derived from applying transposed convolution on coarser scale features and the features from the encoding path to analyze the activations and contextual information for selecting spatial regions.

layer.

$$k = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix} \tag{10}$$

While sharp UNet shows a vast potential for image segmentation, it also comes with certain limitations. The risk of over-sharpening might lead to noise amplification or artifact introduction.

**3.2.2 Attention UNet.** The attention mechanism enhances important features in neural network models. In [42], a new UNet algorithm with an attention module is proposed for pancreas segmentation in CT images. The network includes an encoder module for feature extraction, an attention module to capture contextual information, and a decoder module to restore the concatenated feature maps. Recent studies have shown that combining attention gates (AGs) in DL models can enhance network performance [67, 68] In our study, we present an AG architecture, as depicted in Fig 3, which draws inspiration from the attention gate used in a previous work [42].

The AG unit in Fig 3 acts as a transitional component in the model architecture. It receives the inputs of the decoder and encoder branches. In the encoder branch, convolutional layers employ a hierarchical approach to extract high-level image features by processing local information layer by layer. This process results in the spatial separation of pixels based on their semantics in a higher-dimensional space. By sequentially processing local features, the model can incorporate information from a larger receptive field into its predictions. The feature-map $x_i^l$ represents the extracted high-level features from layer $l$ with the subscript $i$ denoting the spatial dimension for each pixel's value ($a_i$, $x_i$, and $g_i$ in Fig 3). The AG mechanism can be expressed in Eqs 11, and 12.

$$x_{\text{out}} = x_i^l * a_i \qquad (11)$$

where $x_{\text{out}}$ is the feature map calculated by element-wise multiplication between the input feature map $x_i^l$ and attention coefficients $a_i$, which $a_i$ is expressed mathematically as:

$$a_i = \sigma_2(W_s^T(\sigma_1(W_x^T x_i^l + W_g^T g_i + b_g)) + b_s) \qquad (12)$$

where $\sigma_1$ is the swish activation function and $\sigma_2$ is the sigmoid function. The feature map $x_i^l$ and the gating signal vector $g_i$ undergo linear transformations using 1*1 channel-wise convolutions. The parameters $W_x$, $W_g$, and $W_s$ are trainable, and the bias terms $b_g$ and $b_s$ are set to zero for simplicity. Experimental results suggest that setting these bias values to zero does not adversely affect the model's performance.

**3.2.3 The proposed Sharp Attention. UNet.** In this section, we develop a novel architecture by combining the Sharp UNet and Attention UNet models introduced earlier. The motivation behind this fusion is to utilize the advantages of both architectures and extract features that benefit from both the sharpening technique and the attention mechanism. The attention gate features, obtained from the Attention UNet, and the sharpened features derived from the Sharp UNet, are concatenated with the decoder features. This approach aims to enhance the network's capability to capture fine details and relevant contextual information, thus improving the overall feature representation and segmentation performance. The proposed architecture is depicted in Fig 4. In this architecture, we employ the identical sharpening filter and attention gate (AG) mechanism as those introduced in the previously described Sharp UNet and Attention UNet architectures.

## 3.3 Statistical analysis

Performance of each proposed model was evaluated by accuracy, Dice coefficient, loss, Dice loss, precision, sensitivity, specificity, F1, recall, and the Jaccard index. These metrics are explained in detail in S1 Appendix. Performance was compared pairwise between models using the McNemar's test [69], which was performed on a pixel-wise basis for the entire test set comprising 78 images. For each image, the number of discordant entries was calculated by comparing the segmentation results of each two models while keeping the ground truth mask as the true value. Resultant p values were adjusted for multiple comparisons using the Bonferroni correction [70, 71]. P values less than 0.02 were considered significant. All statistical analysis was conducted by using scipy and sklearn packages in Python.

## 4. Results

We randomly divided the BUSI dataset into training, validation, and test sets, with proportions of 80% (624 images), 10% (78 images), and 10% (78 images) respectively. The training set is used to optimize the model's parameters by adjusting them based on input data and
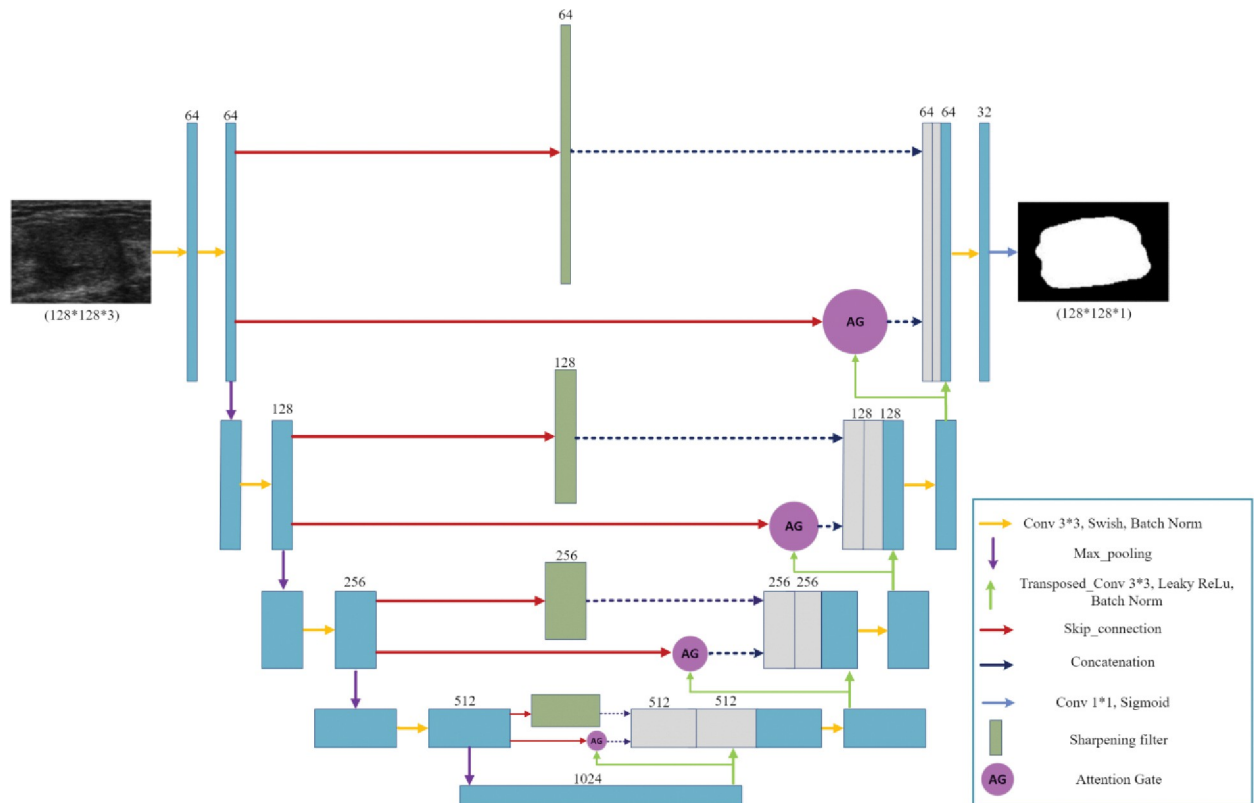
**Fig 4. The proposed Sharp Attention UNet architecture.** The combination of features enhances the network's ability to capture details and context, leading to improved feature representation and segmentation performance.

corresponding target values. After each epoch, the model's performance is evaluated on the validation set to fine-tune hyperparameters such as learning rate, layer count, and neuron count per layer. This process, known as hyperparameter tuning, aimed to optimize performance. The best model is decided based on the validation results and saved as the optimal model. To evaluate performance, a separate test set that was not part of the training or validation processes is utilized. We tested the performance of our proposed Sharp Attention UNet with different input resolutions, as well as an analysis of various activation functions. Furthermore, the impact of applying CLAHE as a preprocessing step is explored. In addition, the proposed model's performance is compared to other models presented in this paper, which were evaluated using the same dataset.

## 4.1 Comparative analysis of neural network models: Performance evaluation and advantages of the proposed architecture

Table 1 displays the performance metrics achieved by the six distinct neural network models presented in this study. The UNet, Attention UNet, and Sharp UNet are built upon previous works ([39, 41, 42] respectively) but fine-tuned and optimized by the hyperparameters presented in this paper. While the primary network architectures for these three models have been previously introduced in existing literature, the same backbone architecture and hyperparameters were used for all models; these will be discussed in detail. For further comparison, we also trained UNet ++ [43], and UNet3+ [44] with our dataset. The image input size of 128*128 was standardized, along with the application of augmentation techniques and

**Table 1. Comparison of the performance of the Sharp Attention UNet model with other UNet based models on the BUSI dataset.**

| Model | Accuracy | Loss | Dice loss | Precision | F1 | Sensitivity | Specificity | Dice coefficient | Jaccard Index |
|---|---|---|---|---|---|---|---|---|---|
| UNet | 0.982 | 0.231 | 0.139 | 0.911 | 0.867 | 0.834 | 0.994 | 0.861 | 0.756 |
| Attention UNet | 0.984 | 0.166 | 0.106 | 0.924 | 0.899 | 0.878 | 0.994 | 0.894 | 0.808 |
| Sharp UNet | **0.984** | 0.186 | 0.123 | 0.885 | 0.886 | 0.890 | 0.992 | 0.877 | 0.781 |
| UNet++ [43] | 0.978 | 0.310 | 0.153 | 0.864 | 0.854 | 0.851 | 0.987 | 0.847 | 0.772 |
| UNet3+ [44] | 0.981 | 0.265 | 0.135 | 0.896 | 0.875+ | 0.858 | 0.992 | 0.865 | 0.775 |
| Proposed model | 0.979 | **0.109** | **0.072** | **0.941** | **0.941** | **0.942** | **0.994** | **0.928** | **0.866** |

https://doi.org/10.1371/journal.pone.0289195.t001

CLAHE preprocessing. The optimization technique, batch size, loss function, and learning rate were kept consistent for all models. The results of these comparisons revealed that the proposed architecture outperformed the other models across almost all validation parameters.

The outcomes achieved by the suggested algorithm indicate enhanced sensitivity, with a notable improvement of more than 5% compared to the second-ranked model. Moreover, the Dice coefficient also experienced a noticeable increase of almost 3% compared to the second-best model. Through experimental evaluations, we demonstrate in Table 1 that our model also outperforms existing state-of-the-art methods in accuracy, robustness, and computational efficiency.

We also applied McNemar's statistical test to compare the dichotomous performance of each pairwise segmentation model evaluated in this paper. McNemar's test can evaluate the statistical significance of differences between the segmentation models [69, 72] under reasonable assumptions. The obtained McNemar's test values are compared to the chi-squared distribution with 1 degree of freedom. The resulting p-value was calculated to determine the statistical significance of the observed differences between the models. The commonly used alpha value is considered 0.05 (5%), indicating that a p-value below 0.05 was considered significant. However, as our database is small, we considered 90% confidence level ($\alpha = 0.1$). Nonetheless, when conducting multiple statistical tests simultaneously, such as comparing multiple pairs of models, there is an increased chance of obtaining false positives. To address this, the Bonferroni correction adjusts the alpha value to maintain an appropriate level of significance [70, 71]. As shown in Table 2, the comparison between the Attention UNet and Sharp UNet models yielded the highest p-value, suggesting a lack of statistically significant distinction between these two models in terms of performance. Our proposed model, which incorporates both Attention gates and sharpening filters within a UNet framework, demonstrated a significant improvement when compared to the earlier models. However, with the considered alpha value in this study, there is not a significant difference between the proposed model and Attention UNet designed in this work. Still, on the parametric evaluation tests, the proposed model shows better results. This finding underscores the synergistic effects of integrating attention mechanisms and sharpening filters, resulting in enhanced performance, and indicating the added value of our proposed model over the standalone Attention UNet and Sharp UNet models.

**Table 2. P values of the McNemar's test for comparing model evaluation.** Values under 0.02 imply the error distribution from the two compared models are significantly different after Bonferonni correction.

| Model | UNet | Attention UNet | Sharp UNet |
|---|---|---|---|
| UNet | | | |
| Attention UNet | **<0.02** | | |
| Sharp UNet | 0.045 | 0.079 | |
| Proposed model | **<0.02** | 0.023 | **<0.02** |

https://doi.org/10.1371/journal.pone.0289195.t002

**Table 3. Quantitative evaluation of Sharp Attention UNet of different input sizes on the BUSI dataset.**

| Input resolution | Batch_size | Accuracy | Loss | Dice loss | Precision | F1 | Sensitivity | Specificity | Dice coefficient | Jaccard Index |
|---|---|---|---|---|---|---|---|---|---|---|
| 32*32 | 32 | 0.974 | 0.274 | 0.138 | 0.908 | 0.868 | 0.834 | 0.991 | 0.862 | 0.757 |
| 64*64 | 32 | **0.980** | 0.188 | 0.096 | 0.923 | 0.909 | 0.899 | **0.994** | 0.904 | 0.824 |
| 128*128 | 32 | 0.979 | **0.109** | **0.072** | **0.941** | **0.941** | **0.943** | 0.994 | **0.928** | **0.866** |

## 4.2 Quantitative analysis of the proposed Sharp Attention UNet model for different input image resolutions

Segmentation was performed using different network input sizes (32*32, 64*64, and 128*128) for our proposed Sharp Attention UNet. However, due to excessive memory usage by the GPU, attempts to increase image resolution higher than 128*128, when the batch size is 32 or greater were not possible. The experiment was performed for 300 training epochs with a batch size of 32. CLAHE was applied as a preprocessing step. Analysis of the performance metrics in Table 3 reveals that the input size of 128*128 achieves the highest values. Hence, we've established the default dimension as 128*128.

## 4.3 Quantitative analysis of the proposed Sharp Attention UNet model for different activation functions

The quantitative analysis conducted on the Sharp Attention UNet model with various activation functions (ReLu, LRelu, swish, and mish) is presented in Table 4. The results indicate that swish outperforms the other activation functions. This suggests that the swish activation function is more suitable for the Sharp Attention UNet model compared to the alternative activation functions that were evaluated.

## 4.4 Quantitative analysis of the proposed Sharp Attention UNet model before and after CLAHE preprocessing

Fig 5 presents a visual representation of two breast images before (a) and after (b) CLAHE enhancement. The samples depicted in this figure comprise two samples from the BUSI dataset. By comparing these images, we can observe the extent to which the CLAHE method enhances the contrast of the images from both datasets.

To assess the effectiveness of CLAHE as the preprocessing technique, the proposed Sharp Attention UNet model was evaluated on the test set using an image size of 128*128 and a batch size of 32. The swish activation function was employed as the activation function. The evaluation aimed to determine the model's efficacy by assessing both the original and CLAHE-enhanced versions of the BUSI dataset. The findings presented in Table 5 demonstrate that the application of CLAHE has a positive effect on the results.

Fig 6 illustrates the contrast between the proposed segmentation model's output and the ground truth. This visual representation offers valuable understanding of the effectiveness and

**Table 4. Quantitative evaluation of the Sharp Attention UNet with different activation functions.**

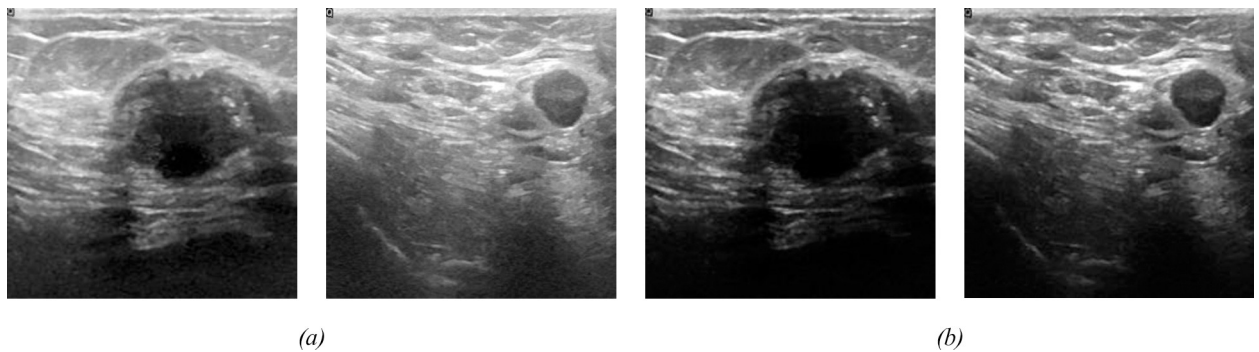| Activation Function | Accuracy | Loss | Dice loss | Precision | F1 | Sensitivity | Specificity | Dice coefficient | Jaccard Index |
|---|---|---|---|---|---|---|---|---|---|
| ReLu | 0.979 | 0.208 | 0.104 | 0.916 | 0.901 | 0.888 | 0.991 | 0.896 | 0.811 |
| LRelu(a = 0.1) | **0.981** | 0.174 | 0.096 | 0.933 | 0.910 | 0.891 | **0.994** | 0.904 | 0.824 |
| Swish | 0.979 | **0.109** | **0.072** | **0.941** | **0.941** | **0.943** | 0.994 | **0.928** | **0.866** |
| Mish | 0.972 | 0.123 | 0.090 | 0.912 | 0.920 | 0.912 | 0.982 | 0.910 | 0.836 |

*(a)* *(b)*

**Fig 5.** (a) Original image (b) Adaptive contrast image from CLAHE.

precision of the proposed model's performance. The generated mask outcomes result from applying a threshold of 0.4 to the predicted mask and converting it into binary form. This threshold value is obtained by applying ROC curve on the processed mask.

## 5. Discussion

In this paper we fine-tuned and evaluated the UNet, Sharp UNet, and Attention UNet models for breast ultrasound segmentation. These networks were chosen due to their high performance in segmenting lesions in medical images. These models, which are already established in the literature, were further optimized and fine-tuned using various techniques to enhance their performance metrics. The application of optimization techniques plays a crucial role in improving the segmentation accuracy of the CNN-based models. By carefully selecting and fine-tuning these techniques, we observed notable improvements in the performance metrics.

Our study demonstrates the efficacy of the proposed Sharp Attention UNet, evidenced by strong performance metrics (dice coefficient, specificity, sensitivity, and F1 score) of 0.93, 0.99, 0.94, and 0.94, respectively. McNemar's test confirmed its significant improvement over UNet and Sharp UNet. Key findings include enhanced dice and sensitivity due to image enhancement (5% and 3%). Moreover, higher resolution (128*128 as compared to 32*32, and 64*64) bolstered dice and sensitivity. While hardware limitations affect its potential for larger resolutions than 128*128 with batch size greater than 32. We also explored various activation functions' impact on segmentation performance. Swish exhibited superior performance to the second-best choice, mish activation function, showing a 3% increase in sensitivity and a 2% enhancement in Dice coefficient.

We explored the combination of Sharp UNet and Attention UNet to create a novel model, known as Sharp Attention UNet. This hybrid architecture demonstrated superior performance compared to the individual models, emphasizing the potential benefits of integrating different network components to increase their respective performance. The advantages of the introduced model over the previous similar architectures are its ability to extract more meaningful features by incorporating sharpening filter and attention gate module. Sharp Attention UNet has several possible uses in clinical practice. It could both potentially replace an experienced

**Table 5. Quantitative evaluation of applying CLAHE to the Sharp Attention UNet.**

| Dataset | Accuracy | Loss | Dice loss | Precision | F1 | Sensitivity | Specificity | Dice coefficient | Jaccard Index |
|---|---|---|---|---|---|---|---|---|---|
| **Original BUSI** | **0.980** | 0.213 | 0.107 | 0.909 | 0.899 | 0.892 | 0.991 | 0.894 | 0.808 |
| **Enhanced BUSI** | 0.979 | **0.109** | **0.072** | **0.941** | **0.941** | **0.943** | **0.994** | **0.928** | **0.866** |

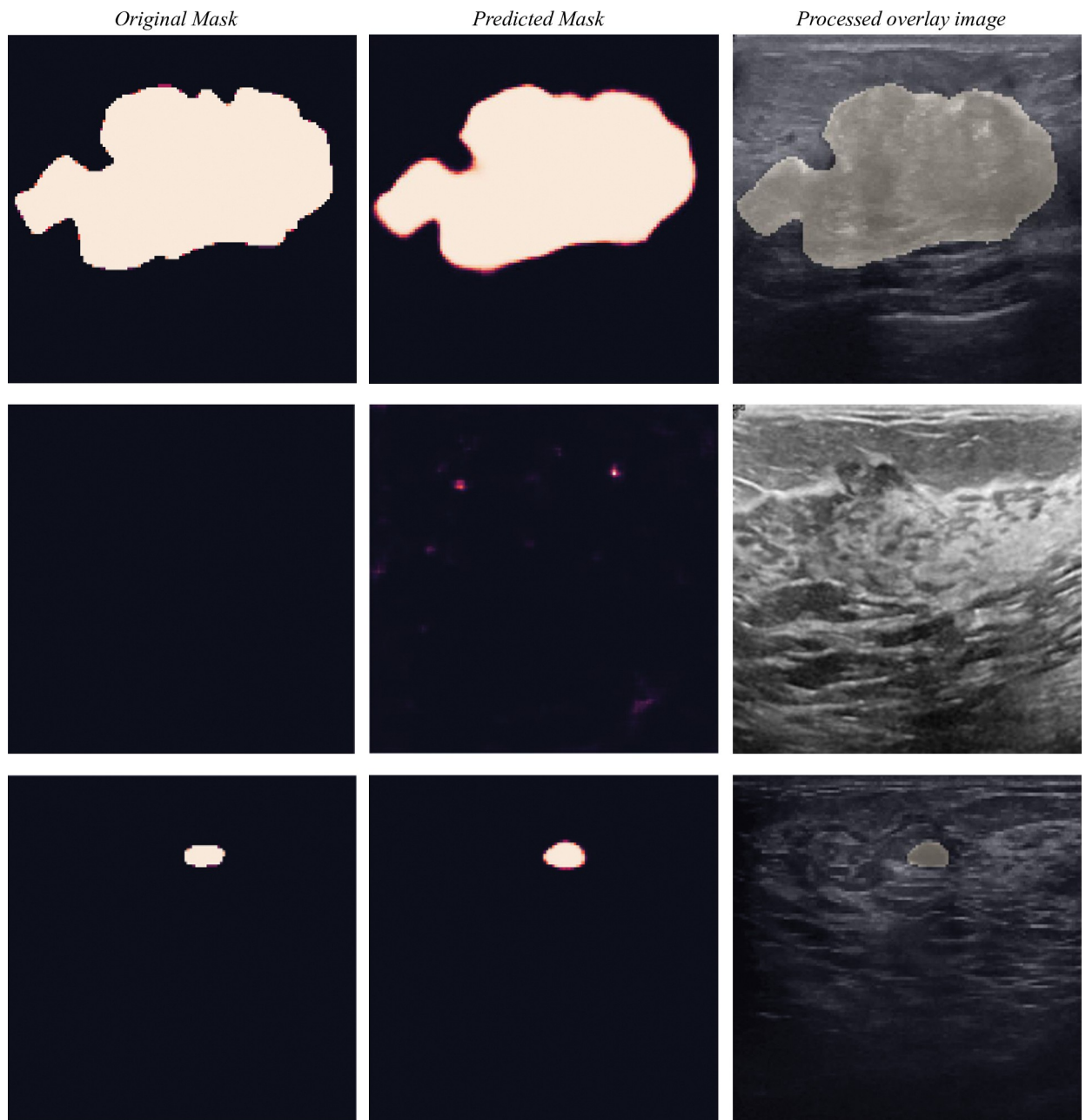|  Original Mask | Predicted Mask | Processed overlay image |

**Fig 6. Some examples of BUSI dataset with their corresponding ground truth, predicted mask, and processed mask overlaid on the original images.** Top row: a malignant lesion, second row: no lesion present, third row: a small benign lesion.

radiologist when one is not available or assist a radiologist to improve diagnostic accuracy. There is also a powerful potential to combine Sharp Attention UNet with standardized image acquisition to allow rapid, automatic diagnosis without a radiologist or sonographer. One such approach incorporates the use of AI with volume sweep imaging (VSI). VSI is an imaging technique in which an individual without prior ultrasound training performs blind sweeps of the ultrasound probe over a target region such as the breast. VSI has already been clinically tested for breast, lung, thyroid, right upper quadrant, and obstetrics indications [6, 9, 73–79]. Individuals have been shown to be able to perform VSI after only a few hours of training [80, 81].

Integration of VSI with AI has already shown promising results for breast and obstetrics indications [12, 82]. Future study testing the performance of Sharp Attention and VSI could be a promising step toward developing rapid and automatic diagnosis of breast lesions without a radiologist or a sonographer.

While the results are promising, it is important to acknowledge the limitations of this study. The evaluation was performed on a specific dataset with a limited number of images, and there may be variations in image quality and characteristics across different datasets. Therefore, further investigation on larger and more diverse datasets is necessary to assess the generalizability and robustness of the proposed models. We will utilize GAN-based data augmentation which has the potential to generate more diverse and realistic synthetic data. Furthermore, the optimization techniques utilized in this work should be studied on other segmentation models.

In future research, it would be also beneficial to explore additional optimization techniques, such as different data augmentation strategies or advanced regularization methods, on different models to further enhance the performance of the segmentation models. Additionally, investigating the transferability of these models to other medical imaging tasks or modalities could expand their potential applications and impact.

## 6. Conclusion

This paper presents a comparative study that explores the impact of different factors such as image preprocessing and various optimization techniques on the performance of UNet, Sharp UNet, and Attention UNet models for breast ultrasound image segmentation. We also introduced a novel UNet-based model, Sharp Attention UNet, by combining the Sharp UNet and Attention UNet models. Sharp Attention UNet could be used to enable rapid automatic diagnosis of breast lesions without a radiologist or sonographer. Since most of the people in the world lack access to any form of medical imaging, this potentially lifesaving artificial intelligence could be a promising avenue to improving global health and the diagnosis of breast cancers.

## Supporting information

**S1 Appendix.**
(DOCX)

## Author Contributions

**Conceptualization:** Donya Khaledyan.

**Data curation:** Donya Khaledyan, Thomas J. Marini, Kevin Parker.

**Formal analysis:** Kevin Parker.

**Funding acquisition:** Kevin Parker.

**Investigation:** Donya Khaledyan, Kevin Parker.

**Methodology:** Donya Khaledyan.

**Project administration:** Kevin Parker.

**Resources:** Kevin Parker.

**Software:** Donya Khaledyan.

**Supervision:** Thomas J. Marini, Kevin Parker.

**Validation:** Donya Khaledyan, Avice O'Connell.

**Visualization:** Donya Khaledyan, Avice O'Connell.

**Writing – original draft:** Donya Khaledyan.

**Writing – review & editing:** Thomas J. Marini, Timothy M. Baran, Avice O'Connell, Kevin Parker.

## References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLO-BOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2018; 68(6):394–424. https://doi.org/10.3322/caac.21492 PMID: 30207593

2. DeSantis CE, Ma J, Gaudet MM, Newman LA, Miller KD, Goding Sauer A, et al. Breast cancer statistics, 2019. CA: a cancer journal for clinicians. 2019; 69(6):438–51. Epub 2019/10/03. https://doi.org/10.3322/caac.21583 PMID: 31577379.

3. O'Connell AM, Marini TJ, Kawakyu-O'Connor DT. Cone-Beam Breast Computed Tomography: Time for a New Paradigm in Breast Imaging. J Clin Med. 2021; 10(21). Epub 2021/11/14. https://doi.org/10.3390/jcm10215135 PMID: 34768656; PubMed Central PMCID: PMC8584471.

4. Tao Z, Shi A, Lu C, Song T, Zhang Z, Zhao J. Breast Cancer: Epidemiology and Etiology. Cell biochemistry and biophysics. 2015; 72(2):333–8. Epub 2014/12/30. https://doi.org/10.1007/s12013-014-0459-6 PMID: 25543329.

5. Tabár L, Vitak B, Chen TH-H, Yen AM-F, Cohen A, Tot T, et al. Swedish Two-County Trial: Impact of Mammographic Screening on Breast Cancer Mortality during 3 Decades. 2011; 260(3):658–63. https://doi.org/10.1148/radiol.11110469 PMID: 21712474.

6. Marini TJ, Oppenheimer DC, Baran TM, Rubens DJ, Toscano M, Drennan K, et al. New Ultrasound Telediagnostic System for Low-Resource Areas. Journal of Ultrasound in Medicine. 2021; 40(3):583–95. https://doi.org/10.1002/jum.15420.

7. Maru DS, Schwarz R, Jason A, Basu S, Sharma A, Moore C. Turning a blind eye: the mobilization of radiology services in resource-poor regions. Global Health. 2010; 6:18. https://doi.org/10.1186/1744-8603-6-18 PMID: 20946643; PubMed Central PMCID: PMC2964530.

8. Ngoya PS, Muhogora WE, Pitcher RD. Defining the diagnostic divide: an analysis of registered radiological equipment resources in a low-income African country. Pan Afr Med J. 2016; 25:99. https://doi.org/10.11604/pamj.2016.25.99.9736 PMID: 28292062; PubMed Central PMCID: PMC5325496.

9. Marini TJ, Castaneda B, Iyer R, Baran TM, Nemer O, Dozier AM, et al. Breast Ultrasound Volume Sweep Imaging: A New Horizon in Expanding Imaging Access for Breast Cancer Detection. J Ultrasound Med. 2022. Epub 2022/07/09. https://doi.org/10.1002/jum.16047 PMID: 35802491.

10. Society AC. Breast cancer facts & figures 2019–2020. Am Cancer Soc. 2019:1–44.

11. Zhang L, Xiao H, Karlan S, Zhou H, Gross J, Elashoff D, et al. Discovery and preclinical validation of salivary transcriptomic and proteomic biomarkers for the non-invasive detection of breast cancer. PloS one. 2010; 5(12):e15573. https://doi.org/10.1371/journal.pone.0015573 PMID: 21217834

12. Marini TJ, Castaneda B, Parker K, Baran TM, Romero S, Iyer R, et al. No sonographer, no radiologist: Assessing accuracy of artificial intelligence on breast ultrasound volume sweep imaging scans. PLOS Digital Health. 2022; 1(11):e0000148. https://doi.org/10.1371/journal.pdig.0000148 PMID: 36812553

13. Boukerroui D, Baskurt A, Noble JA, Basset O. Segmentation of ultrasound images—multiresolution 2D and 3D algorithm based on global and local statistics. Pattern Recognition Letters. 2003; 24(4–5):779–90.

14. Belaid A, Boukerroui D, Maingourd Y, Lerallut J-F. Phase-based level set segmentation of ultrasound images. IEEE Transactions on Information Technology in Biomedicine. 2010; 15(1):138–47.

15. Sarti A, Corsi C, Mazzini E, Lamberti C. Maximum likelihood segmentation of ultrasound images with Rayleigh distribution. IEEE transactions on ultrasonics, ferroelectrics, and frequency control. 2005; 52(6):947–60. https://doi.org/10.1109/tuffc.2005.1504017 PMID: 16118976

16. Su Rehman, Khan MA, Masood A, Almujally NA, Baili J, Alhaisoni M, et al. BRMI-Net: Deep Learning Features and Flower Pollination-Controlled Regula Falsi-Based Feature Selection Framework for Breast Cancer Recognition in Mammography Images. Diagnostics. 2023; 13(9):1618. https://doi.org/10.3390/diagnostics13091618 PMID: 37175009

17. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic

review and meta-analysis. The lancet digital health. 2019; 1(6):e271–e97. https://doi.org/10.1016/S2589-7500(19)30123-2 PMID: 33323251

18. Khaledyan D, Tajally A, Sarkhosh A, Shamsi A, Asgharnezhad H, Khosravi A, et al. Confidence aware neural networks for skin cancer detection. arXiv preprint arXiv:210709118. 2021.

19. Mashhadi N, Khuzani AZ, Heidari M, Khaledyan D, editors. Deep learning denoising for EOG artifacts removal from EEG signals. 2020 IEEE Global Humanitarian Technology Conference (GHTC); 2020: IEEE.

20. Nittas V, Daniore P, Landers C, Gille F, Amann J, Hubbs S, et al. Beyond high hopes: A scoping review of the 2019–2021 scientific discourse on machine learning in medical imaging. PLOS Digital Health. 2023; 2(1):e0000189. https://doi.org/10.1371/journal.pdig.0000189 PMID: 36812620

21. Heidari M, Mirniaharikandehei S, Khuzani AZ, Danala G, Qiu Y, Zheng B. Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms. International journal of medical informatics. 2020; 144:104284. https://doi.org/10.1016/j.ijmedinf.2020.104284 PMID: 32992136

22. Heidari M, Khuzani AZ, Hollingsworth AB, Danala G, Mirniaharikandehei S, Qiu Y, et al. Prediction of breast cancer risk using a machine learning approach embedded with a locality preserving projection algorithm. Physics in Medicine & Biology. 2018; 63(3):035020. https://doi.org/10.1088/1361-6560/aaa1ca PMID: 29239858

23. Fatima M, Khan MA, Shaheen S, Almujally NA, Wang SH. B2C3NetF2: Breast cancer classification using an end-to-end deep learning feature fusion and satin bowerbird optimization controlled Newton Raphson feature selection. CAAI Transactions on Intelligence Technology. 2023.

24. Mashhadi N, Khuzani AZ, Heidari M, Khaledyan D, Teymoori S, editors. Applying a new feature fusion method to classify breast lesions. Medical Imaging 2021: Computer-Aided Diagnosis; 2021: SPIE.

25. Aamir S, Rahim A, Aamir Z, Abbasi SF, Khan MS, Alhaisoni M, et al. Predicting breast cancer leveraging supervised machine learning techniques. Computational and Mathematical Methods in Medicine. 2022; 2022. https://doi.org/10.1155/2022/5869529 PMID: 36017156

26. Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, et al., editors. Swin-unet: Unet-like pure transformer for medical image segmentation. European conference on computer vision; 2022: Springer.

27. Jabeen K, Khan MA, Balili J, Alhaisoni M, Almujally NA, Alrashidi H, et al. BC2NetRF: breast cancer classification from mammogram images using enhanced deep learning features and equilibrium-jaya controlled regula falsi-based features selection. Diagnostics. 2023; 13(7):1238. https://doi.org/10.3390/diagnostics13071238 PMID: 37046456

28. Tan M, Le Q, editors. Efficientnet: Rethinking model scaling for convolutional neural networks. International conference on machine learning; 2019: PMLR.

29. Chaudhury S, Sau K, Khan MA, Shabaz M. Deep transfer learning for IDC breast cancer detection using fast AI technique and Sqeezenet architecture. Math Biosci Eng. 2023; 20:10404–27. https://doi.org/10.3934/mbe.2023457 PMID: 37322939

30. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:160207360. 2016.

31. Bahner DP, Jasne A, Boore S, Mueller A, Cortez E. The ultrasound challenge: a novel approach to medical student ultrasound education. Journal of Ultrasound in Medicine. 2012; 31(12):2013–6. https://doi.org/10.7863/jum.2012.31.12.2013 PMID: 23197555

32. Yang L, Shami A. On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing. 2020; 415:295–316.

33. Sra S, Nowozin S, Wright SJ. Optimization for machine learning: Mit Press; 2012.

34. Sun S, Cao Z, Zhu H, Zhao J. A survey of optimization methods from a machine learning perspective. IEEE transactions on cybernetics. 2019; 50(8):3668–81. https://doi.org/10.1109/TCYB.2019.2950779 PMID: 31751262

35. Ashraf NM, Mostafa RR, Sakr RH, Rashad M. Optimizing hyperparameters of deep reinforcement learning for autonomous driving based on whale optimization algorithm. Plos one. 2021; 16(6):e0252754. https://doi.org/10.1371/journal.pone.0252754 PMID: 34111168

36. Khuzani AZ, Mashhadi N, Heidari M, Khaledyan D, editors. An approach to human iris recognition using quantitative analysis of image features and machine learning. 2020 IEEE Global Humanitarian Technology Conference (GHTC); 2020: IEEE.

37. Probst P, Wright MN, Boulesteix AL. Hyperparameters and tuning strategies for random forest. Wiley Interdisciplinary Reviews: data mining and knowledge discovery. 2019; 9(3):e1301.

38. Hamida S, El Gannour O, Cherradi B, Ouajji H, Raihani A, editors. Optimization of machine learning algorithms hyper-parameters for improving the prediction of patients infected with COVID-19. 2020 ieee

2nd international conference on electronics, control, optimization and computer science (icecocs); 2020: IEEE.

39. Ronneberger O, Fischer P, Brox T, editors. U-net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18; 2015: Springer.

40. Drozdzal M, Vorontsov E, Chartrand G, Kadoury S, Pal C, editors. The importance of skip connections in biomedical image segmentation. International Workshop on Deep Learning in Medical Image Analysis, International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis; 2016: Springer.

41. Zunair H, Hamza AB. Sharp U-Net: Depthwise convolutional network for biomedical image segmentation. Computers in Biology and Medicine. 2021; 136:104699. https://doi.org/10.1016/j.compbiomed.2021.104699 PMID: 34348214

42. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:180403999. 2018.

43. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J, editors. 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, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4; 2018: Springer.

44. Huang H, Lin L, Tong R, Hu H, Zhang Q, Iwamoto Y, et al., editors. Unet 3+: A full-scale connected unet for medical image segmentation. ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2020: IEEE.

45. Yuan Y, Li C, Xu L, Zhu S, Hua Y, Zhang J. CSM-Net: Automatic joint segmentation of intima-media complex and lumen in carotid artery ultrasound images. Computers in Biology and Medicine. 2022; 150:106119. https://doi.org/10.1016/j.compbiomed.2022.106119 PMID: 37859275

46. Chen G-P, Zhao Y, Dai Y, Zhang J-X, Yin X-T, Cui L, et al. Asymmetric U-shaped network with hybrid attention mechanism for kidney ultrasound images segmentation. Expert Systems with Applications. 2023; 212:118847.

47. Byra M, Jarosik P, Szubert A, Galperin M, Ojeda-Fournier H, Olson L, et al. Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network. Biomedical Signal Processing and Control. 2020; 61:102027. https://doi.org/10.1016/j.bspc.2020.102027 PMID: 34703489

48. Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. Dataset of breast ultrasound images. Data in brief. 2020; 28:104863. https://doi.org/10.1016/j.dib.2019.104863 PMID: 31867417

49. Ezatian R, Khaledyan D, Jafari K, Heidari M, Khuzani AZ, Mashhadi N, editors. Image quality enhancement in wireless capsule endoscopy with adaptive fraction gamma transformation and unsharp masking filter. 2020 IEEE Global Humanitarian Technology Conference (GHTC); 2020: IEEE.

50. Razzak MI, Naz S, Zaib A. Deep learning for medical image processing: Overview, challenges and the future. Classification in BioApps: Automation of Decision Making. 2018:323–50.

51. Khaledyan D, Amirany A, Jafari K, Moaiyeri MH, Khuzani AZ, Mashhadi N, editors. Low-cost implementation of bilinear and bicubic image interpolation for real-time image super-resolution. 2020 IEEE Global Humanitarian Technology Conference (GHTC); 2020: IEEE.

52. Khaledyan D, Eshghi M, Heidari M, Khuzani AZ, Mashhadi N, editors. A Practical Method for Pupil segmentation in challenging conditions. 2020 IEEE Global Humanitarian Technology Conference (GHTC); 2020: IEEE.

53. Zuiderveld K. Contrast limited adaptive histogram equalization. Graphics gems. 1994:474–85.

54. Pizer SM, Amburn EP, Austin JD, Cromartie R, Geselowitz A, Greer T, et al. Adaptive histogram equalization and its variations. Computer vision, graphics, and image processing. 1987; 39(3):355–68.

55. Hindi A, Peterson C, Barr RG. Artifacts in diagnostic ultrasound. Reports in Medical Imaging. 2013:29–48.

56. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. Journal of big data. 2019; 6(1):1–48.

57. Yang S, Xiao W, Zhang M, Guo S, Zhao J, Shen F. Image data augmentation for deep learning: A survey. arXiv preprint arXiv:220408610. 2022.

58. Hara K, Saito D, Shouno H, editors. Analysis of function of rectified linear unit used in deep learning. 2015 international joint conference on neural networks (IJCNN); 2015: IEEE.

59. Prajit Ramachandran BZ, Quoc V. Le. Swish: a Self-Gated Activation Function. In International Conference on Learning Representations. ICLR 2017.

60. Misra D. Mish: A self regularized non-monotonic activation function. arXiv preprint arXiv:190808681. 2019.

61. Roelofs R, Shankar V, Recht B, Fridovich-Keil S, Hardt M, Miller J, et al. A meta-analysis of overfitting in machine learning. Advances in Neural Information Processing Systems. 2019; 32.

62. Ying X, editor An overview of overfitting and its solutions. Journal of physics: Conference series; 2019: IOP Publishing.

63. Gal Y, Ghahramani Z, editors. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. international conference on machine learning; 2016: PMLR.

64. Jadon S, editor A survey of loss functions for semantic segmentation. 2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB); 2020: IEEE.

65. Siddique N, Paheding S, Elkin CP, Devabhaktuni V. U-net and its variants for medical image segmentation: A review of theory and applications. Ieee Access. 2021; 9:82031–57.

66. Tan HH, Lim KH, editors. Vanishing gradient mitigation with deep learning neural network optimization. 2019 7th international conference on smart computing & communications (ICSCC); 2019: IEEE.

67. Brauwers G, Frasincar F. A general survey on attention mechanisms in deep learning. IEEE Transactions on Knowledge and Data Engineering. 2021.

68. Hafiz AM, Parah SA, Bhat RUA. Attention mechanisms and deep learning for machine vision: A survey of the state of the art. arXiv preprint arXiv:210607550. 2021.

69. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika. 1947; 12(2):153–7. https://doi.org/10.1007/BF02295996 PMID: 20254758

70. Dunn OJ. Multiple comparisons among means. Journal of the American statistical association. 1961; 56 (293):52–64.

71. Rupert G. Jr Simultaneous statistical inference. 2012.

72. Hawass N. Comparing the sensitivities and specificities of two diagnostic procedures performed on the same group of patients. The British journal of radiology. 1997; 70(832):360–6. https://doi.org/10.1259/bjr.70.832.9166071 PMID: 9166071

73. Marini TJ, Castaneda B, Satheesh M, Zhao YT, Reátegui-Rivera CM, Sifuentes W, et al. Sustainable volume sweep imaging lung teleultrasound in Peru: Public health perspectives from a new frontier in expanding access to imaging. Frontiers in health services. 2023; 3:1002208. Epub 2023/04/20. https://doi.org/10.3389/frhs.2023.1002208 PMID: 37077694; PubMed Central PMCID: PMC10106710.

74. Marini TJ, Kaproth-Joslin K, Ambrosini R, Baran TM, Dozier AM, Zhao YT, et al. Volume sweep imaging lung teleultrasound for detection of COVID-19 in Peru: a multicentre pilot study. 2022; 12(10):e061332. https://doi.org/10.1136/bmjopen-2022-061332%JBMJOpen

75. Marini TJ, Oppenheimer DC, Baran TM, Rubens DJ, Dozier A, Garra B, et al. Testing telediagnostic right upper quadrant abdominal ultrasound in Peru: A new horizon in expanding access to imaging in rural and underserved areas. PloS one. 2021; 16(8):e0255919. Epub 2021/08/12. https://doi.org/10.1371/journal.pone.0255919 PMID: 34379679; PubMed Central PMCID: PMC8357175

76. Marini TJ, Weis JM, Baran TM, Kan J, Meng S, Yeo A, et al. Lung ultrasound volume sweep imaging for respiratory illness: a new horizon in expanding imaging access. BMJ open respiratory research. 2021; 8 (1). Epub 2021/11/14. https://doi.org/10.1136/bmjresp-2021-000919 PMID: 34772730; PubMed Central PMCID: PMC8593737.

77. Marini TJ, Weiss SL, Gupta A, Zhao YT, Baran TM, Garra B, et al. Testing telediagnostic thyroid ultrasound in Peru: a new horizon in expanding access to imaging in rural and underserved areas. Journal of Endocrinological Investigation. 2021. https://doi.org/10.1007/s40618-021-01584-7 PMID: 33970434

78. Toscano M, Marini T, Lennon C, Erlick M, Silva H, Crofton K, et al. Diagnosis of Pregnancy Complications Using Blind Ultrasound Sweeps Performed by Individuals Without Prior Formal Ultrasound Training. Obstetrics & Gynecology. 2023; 141(5). https://doi.org/10.1097/AOG.0000000000005139 PMID: 37103534

79. Toscano M, Marini TJ, Drennan K, Baran TM, Kan J, Garra B, et al. Testing telediagnostic obstetric ultrasound in Peru: a new horizon in expanding access to prenatal ultrasound. BMC Pregnancy and Childbirth. 2021; 21(1):328. https://doi.org/10.1186/s12884-021-03720-w PMID: 33902496

80. Erlick M, Marini T, Drennan K, Dozier A, Castaneda B, Baran T, et al. Assessment of a Brief Standardized Obstetric Ultrasound Training Program for Individuals Without Prior Ultrasound Experience. Ultrasound quarterly. 2022. Epub 2022/10/13. https://doi.org/10.1097/ruq.0000000000000626 PMID: 36223486.

81. Marini T, Castaneda B, Baran T, O'Connor T, Garra B, Tamayo L, et al. Lung Ultrasound Volume Sweep Imaging for Pneumonia Detection in Rural Areas: Piloting Training in Rural Peru. Journal of Clinical Imaging Science. 2019; 9(35). https://doi.org/10.25259/JCIS_29_2019 PMID: 31538033

**82.** Arroyo J, Marini TJ, Saavedra AC, Toscano M, Baran TM, Drennan K, et al. No sonographer, no radiologist: New system for automatic prenatal detection of fetal biometry, fetal presentation, and placental location. PloS one. 2022; 17(2):e0262107. https://doi.org/10.1371/journal.pone.0262107 PMID: 35139093