This paper investigates the application of convolutional neural networks (CNNs), particularly the UNet model architecture, to improve the accuracy of breast cancer tumor segmentation in ultrasound images. Accurate identification of breast cancer is essential for effective patient treatment. However, ultrasound images, often contain noise and artifacts, which can complicate the task of tumor segmentation. Therefore, to highlight the most robust architecture, modifications were made to the original set, including the addition of noise and fuzziness. In this study, a comparative study of five different variants of UNet models (UNet, Attention UNet, UNet++, Dense Inception UNet and Residual UNet) was conducted on a diverse set of ultrasound images with different breast tumors. Using consistent training methods and techniques of augmentation and adding noise to the data, an improvement in segmentation accuracy was highlighted when using the Dense Inception UNet architecture. The results have potential practical applications in the field of medical diagnosis and can assist medical professionals in automatic tumor segmentation in breast cancer ultrasound images. This study highlights the improvement of segmentation accuracy by introducing dense induction into the UNet architecture. Importantly, the Dice coefficient, a key segmentation metric, improved markedly, increasing from 0.973 to 0.976 after data augmentation. The results of the study offer promise to the medical community by offering a more accurate and reliable approach to segmenting breast cancer lesions on ultrasound images. The findings can be implemented in clinical practice to assist radiologists in early cancer diagnosis

Keywords: breast cancer, ultrasound imaging, deep learning, UNet, image segmentation, augmentation

UDC 004.4
DOI: 10.15587/1729-4061.2023.289044

COMPARISON EVALUATION OF UNET-BASED MODELS WITH NOISE AUGMENTATION FOR BREAST CANCER SEGMENTATION ON ULTRASOUND IMAGES

Assel Mukasheva
PhD, Associate Professor
School of Information Technology and Engineering
Kazakh-British Technical University
Tole Bi str., 59, Almaty, Republic of Kazakhstan, 050000
Dina Koishiyeva
Corresponding author
Master's Student
Department of Information Systems and Cybersecurity*
E-mail: d.koishiyeva@aues.kz
Zhanna Suimenbayeva
PhD Doctorate
Department of Cybersecurity, Information Processing and Storage
Satbayev University
Satpayev str., 22a, Almaty, Republic of Kazakhstan, 050013
Sabina Rakhmetulayeva
PhD, Associate Professor**
Aigerim Bolshibayeva
PhD, Assistant Professor**
Gulnar Sadikova
PhD Doctorate
Department of Telecommunications and Space Engineering*
*Almaty University of Power Engineering and Telecommunications
Baytursinuli str., 126/1, Almaty, Republic of Kazakhstan, 050013
**Department of Information Systems
International Information Technology University
Manas str., 34/1, Almaty, Republic of Kazakhstan, 050000

Received date 11.08.2023
Accepted date 19.10.2023
Published date 30.10.2023

How to Cite: *Mukasheva, A., Koishiyeva, D., Suimenbayeva, Z., Rakhmetulayeva, S., Bolshibayeva, A., Sadikova, G.(2023). Comparison evaluation of unet-based models with noise augmentation for breast cancer segmentation on ultrasound images. Eastern-European Journal of Enterprise Technologies, 5 (9 (125)), 85–97. doi: https://doi.org/10.15587/1729-4061.2023. 289044*

1. Introduction

Oncology is a global problem with a significant impact on the health and lives of people around the world. Breast cancer occupies a special place as one of the most frequently diagnosed oncological diseases in women [1]. In Republic of Kazakhstan, breast cancer occupies one of the first places in the list of malignant neoplasms affecting women [2]. According to statistics, about 4.6 thousand new cases of this formidable disease and, unfortunately, 1.3 thousand deaths are registered annually in Republic of Kazakhstan [3]. Early diagnosis plays a crucial role in preventing the spread of cancer cells through the lymphatic pathways and damage to critical organs. Detecting malignant neoplasms in their early stages allows healthcare providers to take timely action to limit the spread of cancer and preserve the integrity of vital body structures. Methods include invasive biopsy and non-invasive techniques such as ultrasound [4]. Breast ultrasound (BUS) is a convenient and valuable diagnostic method available for early detection of breast cancer in all age groups of women [5]. However, the complexity of interpreting these images is attributed to their low reso-

lution and the interference of speckle noise. In response, computer-aided diagnosis (CAD) systems employing deep learning (DL) [6] have been developed to streamline ultrasound image analysis and offer precise decision support for radiologists. For automated breast cancer segmentation using deep learning, it is necessary to select models and algorithms that give the best results [7]. This allows changes to be made to the ultrasound images and allows the tumor boundaries to be accurately defined [8]. The development of an accurate and user-friendly computer-aided design system for accurate tumor identification (malignant or benign) using machine learning has the potential to improve the efficiency of mammologists, improve diagnosis and contain cancer progression [9]. Hopefully, with the help of such systems, it will be possible to reduce the progression of such a terrible disease [10].

2. Literature review and problem statement

This section presents the results of research devoted to the important task of breast cancer diagnostics using ultrasound images. It is shown that early detection of breast cancer is the key to successful treatment and improved patient outcomes. However, despite significant advances in medical imaging, there remain unresolved issues related to the accuracy of tumor delineation on ultrasound images.

Breast cancer is a diverse disease with subtypes such as ductal carcinoma and lobular cancer. Distinguishing between benign and malignant lesions and differentiating the various subtypes of cancer is crucial for clinical decision-making. The reason for this may be the objective difficulties associated with the complex texture and variability of breast tissue. These difficulties make manual segmentation impractical and hinder the development of accurate diagnostic tools. One way to overcome these difficulties may be the application of deep learning techniques. This approach has been used in previous studies, but there is still room for improvement. Despite the extensive use of deep learning in medical imaging tasks, including breast cancer detection, several significant issues remain. One of these is the selection of the most appropriate neural network architecture and optimal hyper parameter tuning. This is particularly important since ultrasound images used in diagnosis are characterized by different types of noise and complex textures, which complicates the segmentation and analysis process. Therefore, the selection of suitable algorithms that meet these requirements becomes imperative. Most of the deep learning models applied to image segmentation problems use some or the other techniques related to the encoder-decoder architecture.

There are two categories of encoder-decoders for solving image segmentation problems – general segmentation and biomedical segmentation approaches. The paper [11] offers a UNet model which belongs to the CNN group, is designed specifically for medical segmentation problems. The UNet model's advantages lie in its encoder-decoder structure and skip connections, which enable feature extraction, image recovery, and the combination of low-level and high-level features for effective segmentation, and UNet network has the disadvantage of a small receptive field and cannot use the information relationship between different channels. Currently, there are many extensions to the basic UNet architecture. Over time, it has become apparent that there is potential to improve this architecture. This literature review lists significant models that utilize UNet as well as additional blocks in their architecture. Along with these techniques, the work [12] developed of particular interest is the UNet++ architecture which is based on UNet layers with path skipping, a key advantage of UNet++ is its improved segmentation performance compared to the traditional UNet architecture, achieved by bridging the semantic gap between encoder and decoder blocks through nested convolutional blocks. The disadvantage of UNet++ is the increased computational complexity due to the additional nested convolutional blocks and skip connections, which can result in higher resource requirements during training and inference. There is another [13] notable approach is the Attention UNet model, which uses "attention gates" to emphasize the most important areas. The advantage of the Attention UNet model is its ability to highlight and pay attention to important parts of images, which can improve segmentation accuracy in complex scenarios. However, the disadvantage of this model can be higher training time and computational resource requirements due to the attention paid to each pixel in the image. The work [14], suggests besides UNet++ and Attention UNet, another technique, Residual UNet can be distinguished. It utilizes residual links derived from the ResNet architecture [15], which allows more efficient transfer of information between different layers of the network and facilitates the training of deep models for segmentation. At the same time, the disadvantage can be higher memory and computational power consumption due to the increased network depth. Moreover, the method proposed in [16] presented the Dense-Inception UNet model which combines U-Net layers and also uses modules – Inception-Res and Dense Connecting module, enhancing feature extraction and segmentation accuracy in medical images, however, due to its complex architecture, it may require more computational resources and result in slightly longer training times. In [17], models specifically designed for breast cancer segmentation in ultrasound images will be discussed. A study that proposed to use models including the DL-UNet model, to address unclear segmentation boundaries, but it may still face challenges in handling intricate image features, especially in scenarios with diverse and complex anatomical structures, further research may be needed to enhance its ability to capture fine-grained details. According to the study [18], the RMS-SE-UNet model performed well on the ultrasound image dataset with a Dice coefficient of 94.69 % including residual block, pass coupling, compression, and excitation (SE) blocks, the model's exceptional performance on one dataset may not necessarily translate to robust performance across various medical imaging datasets with distinct characteristics. The RDAUNet model combines UNet layers with the Wasserstein GAN (WGAN) algorithm. In particular, it achieved overall Accuracy, PR-AUC, ROC-AUC and F1-score of 0.98, 0.95, 0.89 and 0.88, respectively, however, it's worth noting that the utilization of the WGAN algorithm may significantly increase the computational demands, making it less accessible for resource-constrained environments [19]. Similarly, InvUNet model based on UNet architecture has involutional layers, nevertheless, the introduction of involutional layers adds a layer of complexity to the model, which may require careful tuning and increased computational resources during training and inference [20]. Recent research [21] proposes an AAU-Net is a model that adds a hybrid adaptive attention module (HAAM) to the

layers underlying the UNet architecture. However, the incorporation of HAAM might lead to higher computational requirements, necessitating efficient hardware.

In addition, models may require large amounts of training data and parameter tuning to achieve optimal results. To conclude the literature review, there are currently a large number of algorithms based on the UNet architecture, the challenge is to select the most optimal architecture that meets the requirements of breast cancer tumor segmentation on ultrasound images. This study is expected to contribute to more accurate diagnosis and better treatment planning, ultimately leading to improved patient outcomes.

3. The aim and objectives of the study

The aim of this study is to improve the segmentation process of breast cancer tumors on ultrasound images using deep learning to provide a more accurate and efficient diagnosis, which in turn contributes to better treatment of patients.

To achieve the objective, the following objectives were set:

– to perform a comparative evaluation of different image segmentation methods, parti-cularly models based on UNet architecture, to distinguishing between healthy breast tissue and cancerous tumors within ultrasound images;

– to investigate the impact of data augmentation techni-ques, such as introducing noise and fuzziness, on the segmentation accuracy of breast cancer tumors across the evaluated methods.

4. Materials and methods

4. 1. Object and hypothesis of the study

The research object is different models based on UNet architecture for breast tumor segmentation on medical ultrasound images. The main hypothesis of the study is that introducing speckle noise and fuzziness into the training data and then training models on this data can help to determine the most robust and appropriate architecture for this task. The hypothesis aims to test whether data augmentation techniques, such as the introduction of speckle noise and fuzziness, can improve the robustness and performance of UNet-based models in this specific medical imaging task. The simplifications adopted in this study are related to the data augmentation process. Various data augmentation techniques were used, including image rotation, random shifts, shading and the introduction of random Gaussian noise (with a standard deviation of 0.05). These were implemented to improve the model's ability to generalize and deal effectively with variations in medical ultrasound images, ultimately improving segmentation performance. These simplifications are expected to improve the robustness and reliability of the developed models for breast tumor segmentation and provide valuable insights for practitioners in the field of early cancer diagnosis.

4. 2. Architectural model descriptions

U-Net model architecture.

The UNet architecture shown Fig. 1 is U-shaped with two paths, tapering (left) and widening (right) for tumor segment output. The input data of the model is 128×128 pixels. The total number with parameters is 7,771,939; trainable, 7,766,051; untrainable, 5,888.

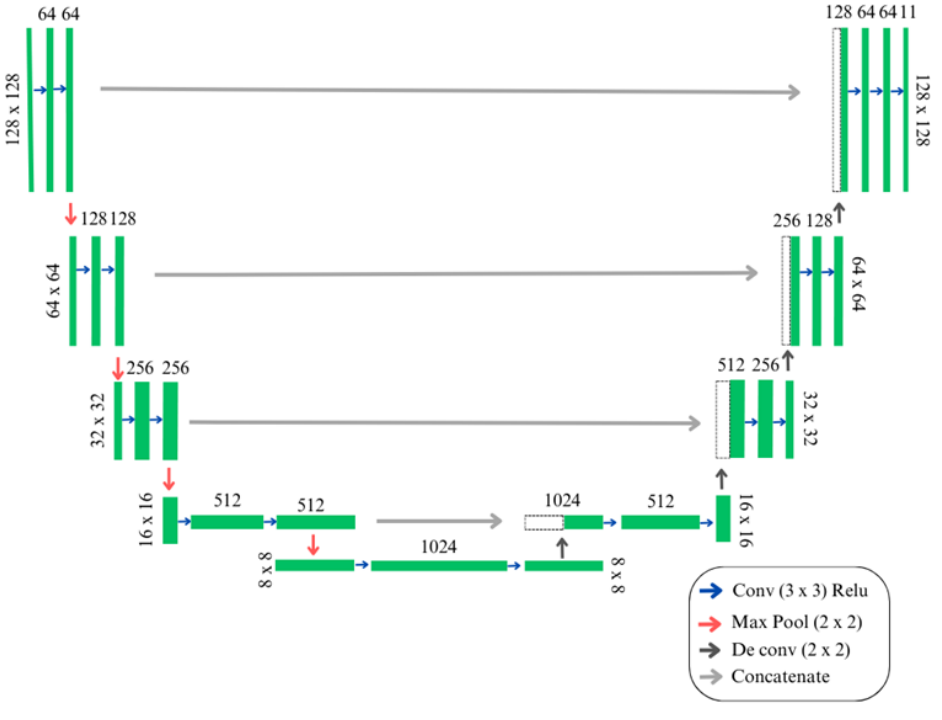

Fig. 1. UNet model architecture

The picture above Fig. 1 shows the layers of the architecture are 3×3 convolution and ReLU [22] activation these are used for data processing and Max-pooling, channel reduction and feature concatenation are used in the expanding path [23]. Model uses an energy function that compares the model segment predictions with the original data. If the model is wrong, the function shows a larger error:

$$E = \Sigma_{x\in\Omega} w(x)\log\left(pl_{(x)}(x)\right), \quad (1)$$

where Ω represents the true label of each pixel. The weight (w) considers the importance of pixels in training the model.

UNet++(Nested) model architecture.

The architecture UNet++ is a deeply supervised encoder-decoder network, with nested, dense skip connections linking the encoder and decoder subnetworks. These connections aim to narrow the semantic gap between their feature maps Fig. 2.

UNet++ improves the relation between encoder and decoder, having in addition to the basic architecture blocks of convolutions depending on the pyramid layer:

$$x^{i,j}=\begin{cases}\mathrm{H}\left(x^{i-1,j}\right), j=0,\\ \mathrm{H}\left(\left[x^{i,k}\right]_{k=0}^{j-1},\cup(xi+1,j-1), j>0.\end{cases} \quad (2)$$

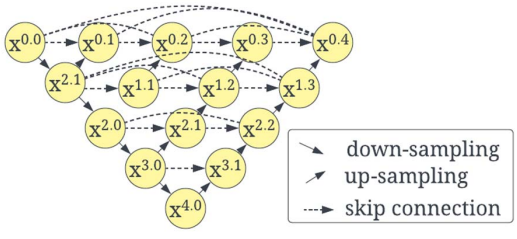

Fig. 2. UNet++(Nested) model architecture

Transition paths with different numbers of input features: this means that there are transition paths in the model that can receive different numbers of input features. In particular, these transition paths receive inputs from different levels of coding layers. The main difference between UNet++ and the basic UNet model is manifested by the fact that transition paths with variable number of inputs are introduced at different layers of the network.

Attention UNet model architecture.

Attention UNet implements attention gates as shown in Fig. 3 below (AGs) in UNet AGs highlight important features and reduce noise by filtering activations both ways.

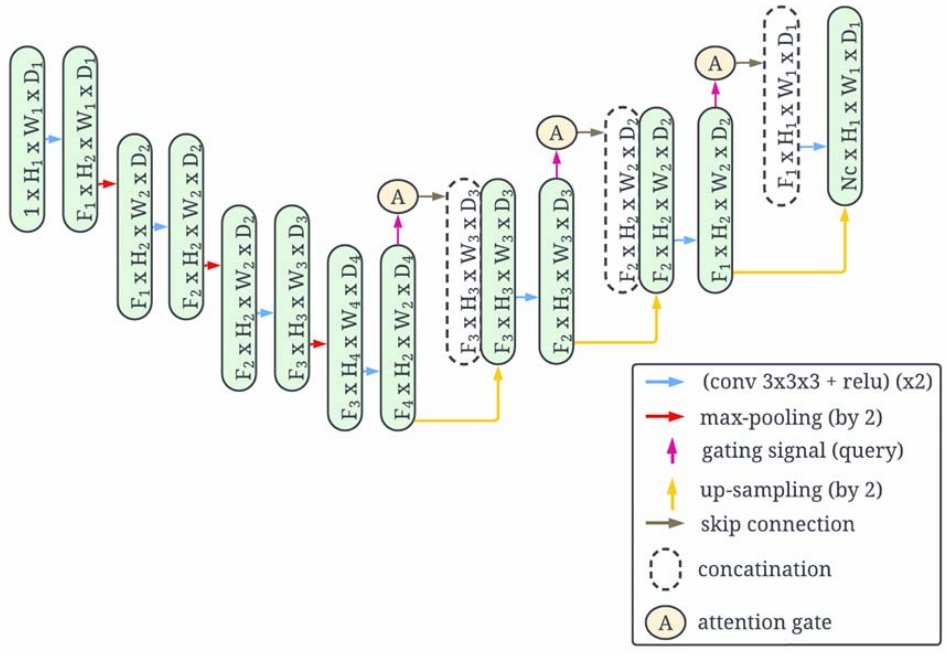

Fig. 3. Attention UNet architecture

An important feature is that AGs contribute to the backpropagation of the error by skipping more relevant gradients and reducing the influence of background regions. As a result, the introduction of the attention mechanism into the UNet architecture contributes to more accurate segmentation.

Residual UNet model architecture.

This model combines the UNet architecture and residual neural networks. By the inclusion of the residual block, learning becomes more controllable. Below is a description of the residual block of Fig. 4, the deepening of which in a multilayer neural network can improve its performance [24].

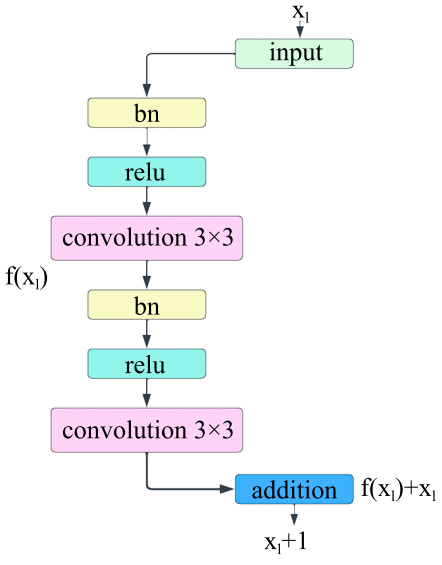

Fig. 4. Residual block

The residual blocks may be shown in the form (4):

$$y_l=h(x_l)+F(x_l w_l),$$
$$x_{l+1}=f(y_l). \quad (3)$$

Residual blocks, as an addition to the main UNet layers, can improve the model's extractions of features of interest.

Dense Inception UNet model architecture.

The model architecture combines the basic UNet layers in addition to the Inception-Res and Dense-Net blocks.

In the above Fig. 5 shows Residual Inception block, where it is possible to see the layers of residual concatenation. By concatenating feature maps from different nodes of different branches with different kernels, which increases the network width and extracts the desired features for segment extraction:

$$x_{l+1}=f_{IR}\left(\left[x_0,x_1,\ldots,x_l\right]\right). \quad (4)$$

In this way, adding the Dense-Inception block to the underlying architecture increases the depth and width of the network and improves its functionality.

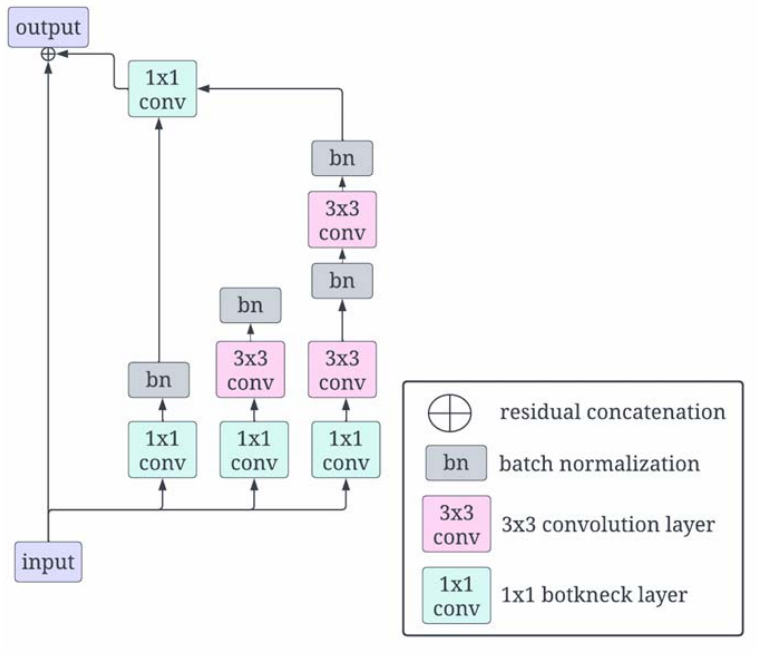

Fig. 5. Inception Residual block

Model evaluation metrics.

In this particular study, metrics namely accuracy, Dice coefficient, F1 score, recall, precision and mean Intersection over Union (IoU) are used to evaluate the performance of the models. A description of these metrics is presented below:

– accuracy. Measures the overall correctness of the prediction: the percentage of correctly classified samples (true positive and true negative) from the total data:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}; \quad (5)$$

– dice index: that metric measures the similarity in image segmentation. It evaluates the consistency between two sets A (segmented regions) and B (benchmark regions):

$$\text{Dice}(A,B) = 2\frac{|\text{A}| \cap |\text{B}|}{|\text{A}| + |\text{B}|}; \quad (6)$$

– F1 Score: metric combines precision and fullness:

$$F1 = 2\frac{\text{precission} * \text{recall}}{\text{precission} + \text{recall}}; \quad (7)$$

– recall: this metric measures the percentage of correctly classified samples, including both True Positives (*TP*) and True Negatives (*TN*):

$$\text{Recall} = \frac{TP}{TP + FN}; \quad (8)$$

– precision: metric assesses the accuracy of positive forecasts by measuring the proportion of TP among all forecasts as positive:

$$\text{Precission} = 100 \times \frac{TP}{TP + FP}; \quad (9)$$

– mean IOU: measures the similarity between the predicted segmentation and the benchmark through intersection and union of masks [25–32]:

$$\text{IoU} = \frac{|Mask \cap prediction|}{Mask \cup prediction}. \quad (10)$$

It is expected that the use of the above metrics will help to evaluate the effectiveness of the UNet model and the efficiency of its segmentation.

4. 3. Dataset description and preprocessing

In this research, the models are trained on a publicly available dataset sourced from Kaggle. The dataset was gathered in 2018 and is centered around breast ultrasound images obtained from women aged 25 to 75. It consists of 600 female patients, comprising 780 images, each with an average size of 500*500 pixels in PNG format [33].

Fig. 6–8 are samples: original breast ultrasound images paired with corresponding masks. The images show the scans and the masks highlight specific areas of interest in each scan.

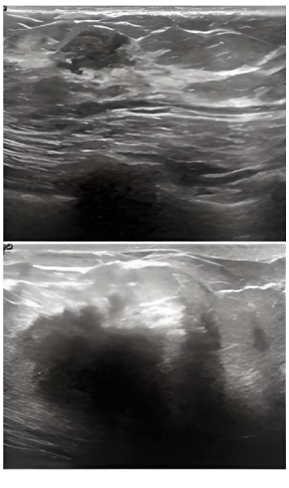

Fig. 6. Sample images from the original dataset, original images

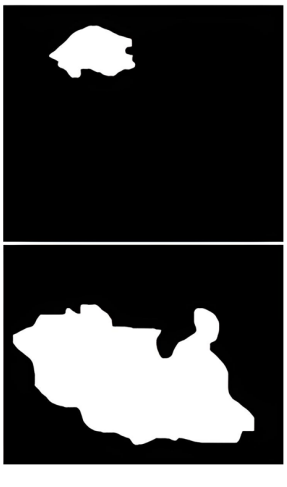

Fig. 7. Sample images from the original dataset, original masks

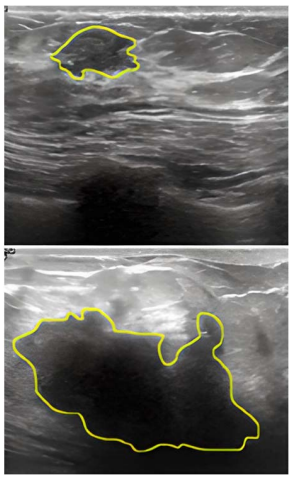

Fig. 8. Sample images from the original dataset, original image with mask contour

Data preprocessing.

To start training, the original dataset consisting of 500×500 pixel images was adapted to a size of 128x128 pixels to fit the architecture of the models. Before training started, the original dataset consisting of 780 samples was divided into two parts. 700 samples were used in the training set and the remaining 80 samples were set aside for validation. After training was completed, evaluation was performed and prediction results were obtained.

The schematic of Fig. 9 illustrates the successive stages of the neural network training process. These steps include uploading data to the server, data preprocessing, training preparation, validation, evaluation and prediction of segment output data.

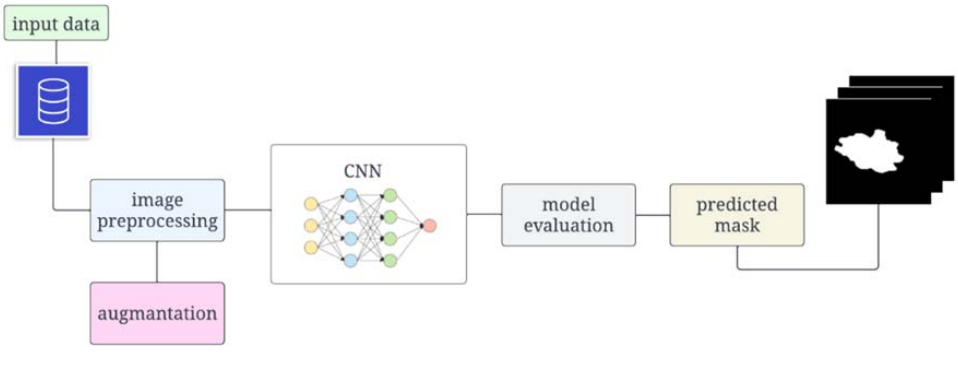

Fig. 9. Flow diagram for Breast Cancer Segmentation

After training the models and analyzing the results, the dataset was increased from 780 to 2000 images. In this process, small rotations, noise, and blurriness were added to the images. The models were then retrained on this expanded dataset. This will be described in detail below.

Data Augmentation.

In this study, a data augmentation strategy was used to increase the versatility of the models. For this purpose, an approach based on the TensorFlow version 2.12.0 and Keras Image Data Generator class was used for automatic image augmentation. The augmentation process included several steps: rotating the image and mask in a given range of angles to train on different orientations (–30 to 30 degrees); random horizontal and vertical shifts of 20 % of the image size; nearest neighbor fill method to avoid resizing artifacts; and adding random Gaussian Noise (0.05) [34]. The original images before modification are shown below in Fig. 10.

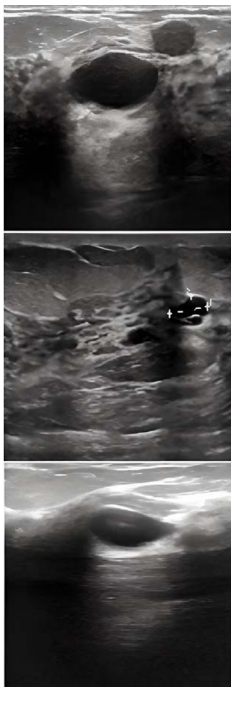

Fig. 10. Sample images original images

Sample images after augmentations with noise are shown in Fig. 11.

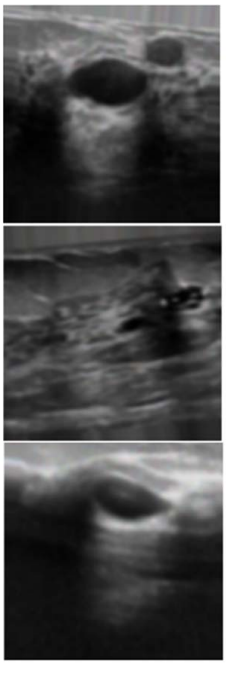

Fig. 11. Sample images after augmentation with noise addition

Hardware and Software Configuration.

The research was conducted using Python 3.6.8 as the programming language. The models were trained on the server with the following specifications: Intel(R) Xeon(R)

Gold 6338 CPU, 512 GB of RAM, and an Nvidia A100 80 GB graphics card.

5. Results of study breast cancer segmentation

5. 1. Comparative assessment of segmentation methods

After completing training and testing on a dataset of 780 original ultrasound images, the models underwent retraining using augmented data. The performance evaluation of segmentation models on breast ultrasound images before data augmentation Table 1 will be summarized below.

The models, trained for a total of 210 epochs, showcase noteworthy results. For instance, Dense Inception UNet demonstrates robust accuracy (0.4234) and Dice Coefficient (0.9734), while Attention UNet excaccuracy Score (0.2067) and Precision (0.9831). UNet++ exhibits a remarkable Mean IoU (0.9578). Additionally, the training progress of these models is depicted through learning curves.

Table 1

Performance evaluation of models on original images

Model	Loss	Acc	Dice	F1	Mean IoU	Precision	Recall
UNet	0.0069	0.9859	0.9642	0.9763	0.9282	0.9649	0.9651
AttUNet	0.43	0.2067	0.9681	0.9832	0.9663	0.9831	0.9834
Unet++	0.58	0.3396	0.9786	0.9788	0.9578	0.9791	0.9787
DIU	0.19	0.4234	0.9734	0.9924	0.9838	0.9926	Nan
ResUnet	0.79	0.1187	0.9909	0.9712	0.9431	0.9698	0.9727

The diagram Fig. 12 illustrates the progression of Dice coefficient values across the training and validation phases for different models. This visualization provides insights into the models' ability to delineate object boundaries and the accuracy of their segmentation performance.

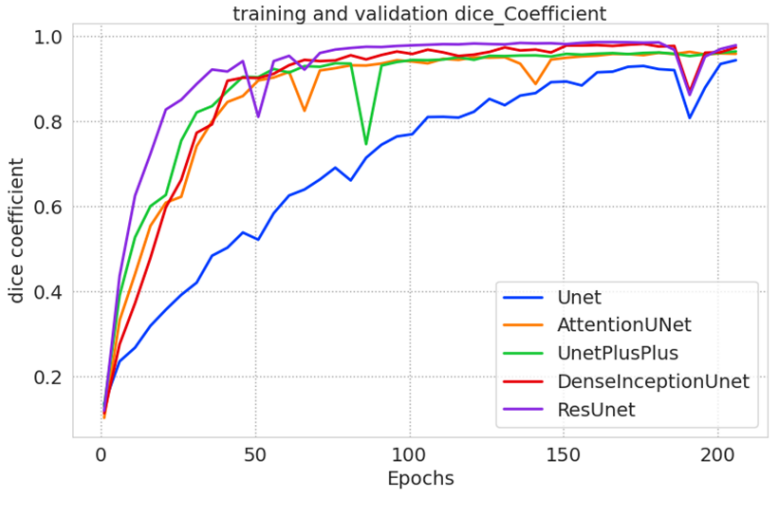

Fig. 12. Training and validation Dice coefficient dynamics model on original images

This visual illustration Fig. 13 allows to see how the loss values change during training and validation at different epochs for different models.

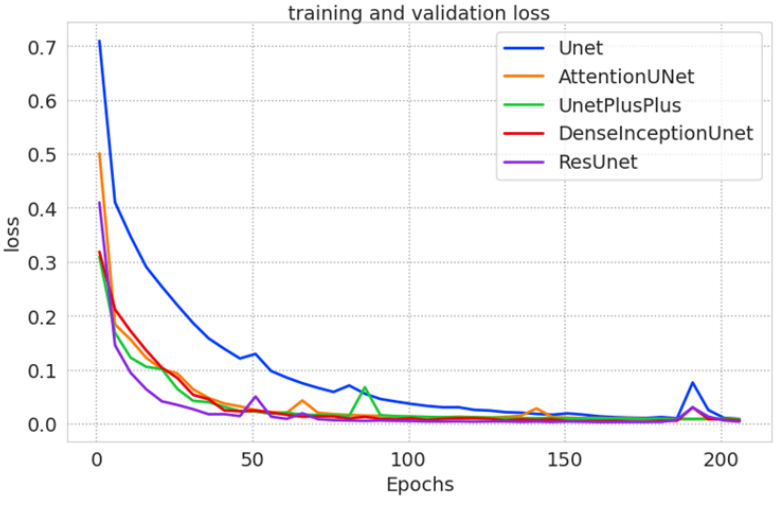

Fig. 13. Training and validation loss coefficient dynamics model on original images

It is worth noting that the Dense Inception UNet model shows the lowest loss value (0.19), indicating efficient optimization and strong generalization. On the other hand, models with relatively high, as demonstrated by examples such as UNet (0.69) and Residual UNet (0.79), the figure unveils notable loss fluctuations. These fluctuations can suggest potential intricacies in fine-tuning specific architectural facets.

5. 2. Effect of data augmentation methods

In this subsection, let's present the results of the evaluation of the model metrics presented in Table 2. These results will provide insight into how much the performance of each model variant improves and whether it is affected by noise and image blurring.

The following conclusions can be drawn about the performance of the models: UNet achieved high values in Precision (0.986), Dice (0.978), and IoU (0.952). Attention UNet excelled in Dice (0.860) and Precision (0.927) despite lower accuracy (0.973). UNet++ demonstrated notable Precision (0.985) and Dice (0.966) scores. Dense Inception UNet and Residual UNet displayed similar results with strong Precision (0.979, 0.986) and Dice (0.977) metrics. Based on the main metrics, Residual UNet and Dense Inception UNet (DIU) models show the best results.

Displayed below are the curves depicting the Dice coefficients and loss metrics from the training history of each model.

The curve Fig. 14 showcases the changes in Dice coefficients over the course of training and validation for all models with augmented images.

Table 2

Performance evaluation results of models after augmentation of the images

Model	Loss	Acc	Dice	F1	Mean IoU	Precision	Recall
UNet	0.065	0.986	0.978	0.977	0.952	0.977	0.977
AttUNet	0.046	0.973	0.860	0.900	0.818	0.927	0.879
UNet++	0.010	0.985	0.966	0.968	0.934	0.969	0.867
DIU	0.007	0.986	0.976	0.977	0.949	0.979	0.974
ResUnet	0.006	0.986	0.977	0.976	0.951	0.976	0.976

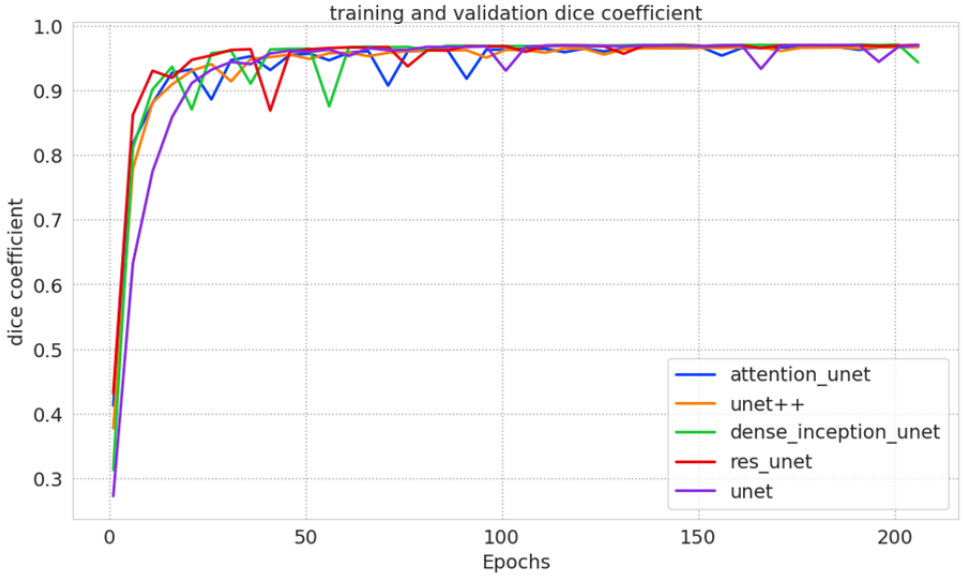

Fig. 14. Comparison of training and validation Dice for all models after augmentation images

In conclusion, the training history graph Fig. 14 provides valuable insights into the models' learning process, primarily through the progression of loss values. By examining the data presented in Table 3, it is possible to discern the varying performance of the models in minimizing prediction errors. The Residual UNet model stands out as the most effective in this regard, achieving an impressively low loss value of 0.006. Similarly, the Dense Inception UNet model showcases strong performance with a minimal loss of 0.007. These findings collectively underline the importance of model architecture and training strategies in optimizing accuracy and performance in medical image analysis. The curve in Fig. 15 shows the comparison loss for all models after increasing the data set.

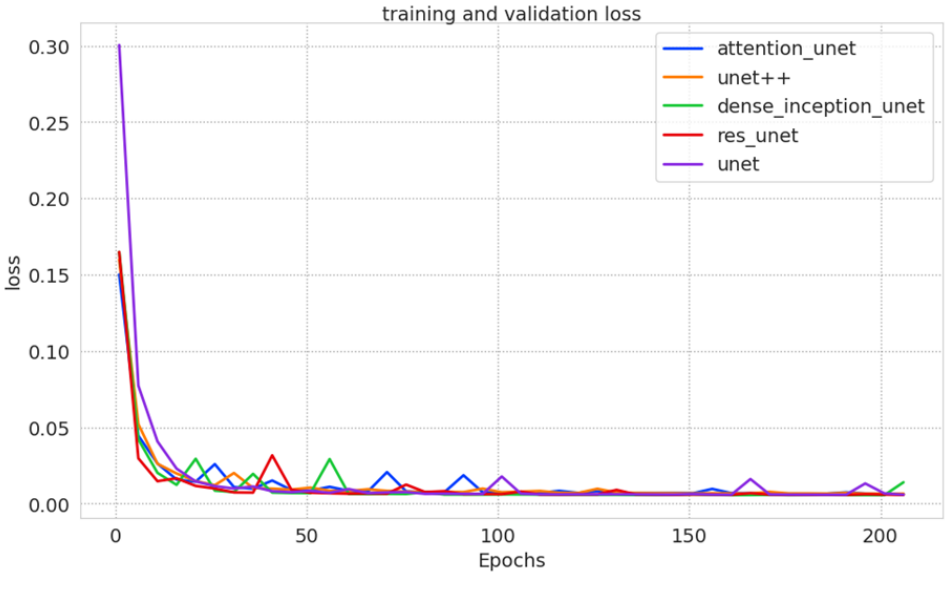

Fig. 15. Comparison Loss for all models after dataset augmentation

Fig. 16 below shows the results of the Dice coefficient metric comparison before and after the introduction of augmentation with noise in percent. In this diagram, the impact of the changes can be clearly compared.

In the comparison chart of Dice scores before and after augmentation, let's assess the performance of five distinct models: UNet, Attention Unet, Unet++, DIU, and Residual Unet, with their Dice scores representing the accuracy of image segmentation, all depicted in percentages.

For UNet, the Dice score improved from 96.42 % before augmentation to 97.80 % after augmentation, indicating a noteworthy 1.38 % increase. In the case of Attention Unet, however, the Dice score experienced a noticeable decline, dropping from 96.81 % before augmentation to 86.00 % after augmentation, resulting in a substantial reduction of 10.81 %. Similarly, the Unet++ model exhibited a slight decrease, shifting from 97.86 % to 96.60 %, signifying a modest 1.26 % drop in the Dice score.

Conversely, the DIU model demonstrated a marginal increase, with the Dice score moving from 97.34 % before augmentation to 97.60 % after augmentation, equating to a minor uptick of 0.26 %. Finally, the Residual Unet model experienced a reduction, moving from 99.09 % before augmentation to 97.70 % after augmentation, reflecting a decrease of 1.39 %.

These alterations in Dice scores before and after augmentation offer valuable insights into how each model's performance responds to the data augmentation process.

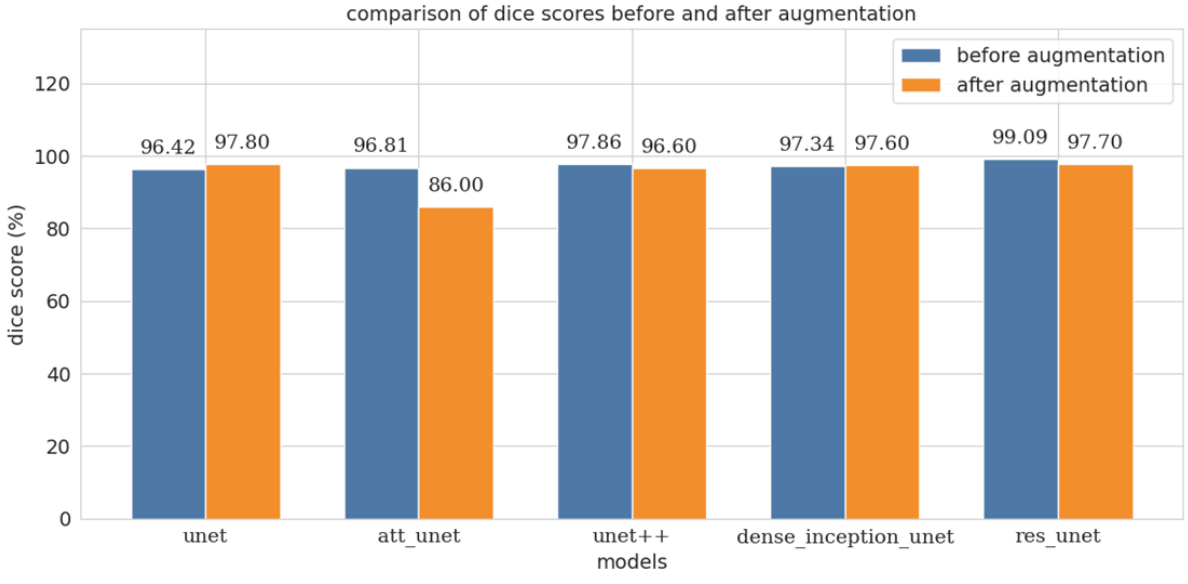

Fig. 16. Dice coefficient for all models before and after dataset augmentation (%)

The following Fig. 17 shows a comparison chart in (%) for the Precision metric comparing before and after the introduction of augmentation.

For UNet, the Precision score improved from 96.49 % before augmentation to 97.70 % after augmentation, indicating a notable 1.21 % increase. In the case of Attention UNet, however, the Precision score experienced a noticeable decline, dropping from 98.31 % before augmentation to 92.70 % after augmentation, resulting in a substantial reduction of 5.61 %. Similarly, the UNet++ model exhibited a slight decrease, shifting from 97.91 % to 96.90 %, signifying a modest 1.01 % drop in the Precision score.

These alterations in Precision scores before and after augmentation offer valuable insights into how each model's performance responds to the data augmentation process, with the effects being slightly less pronounced compared to the Dice scores.

Below Table 3 are the results of the training speed of each model on the original dataset. Values are given in ms and rounded to the nearest integer ms.

Table 3

Training Time Comparison all models on the original dataset

Model	Training time (ms)
UNet	171788
Attention UNet	178499
UNet++	182925
Dense Inception UNet	442597
Residual UNet	190108

From the results of Table 3 training time, it can be seen that the UNet model with dense induction has a high learning rate of 442597 ms on the original dataset, while the UNet model has the lowest training time of 171788 ms among all the models. The results of the training time on the augmented data are as follows. Next, Table 4. below presents the results of the training time, after the change in the data.

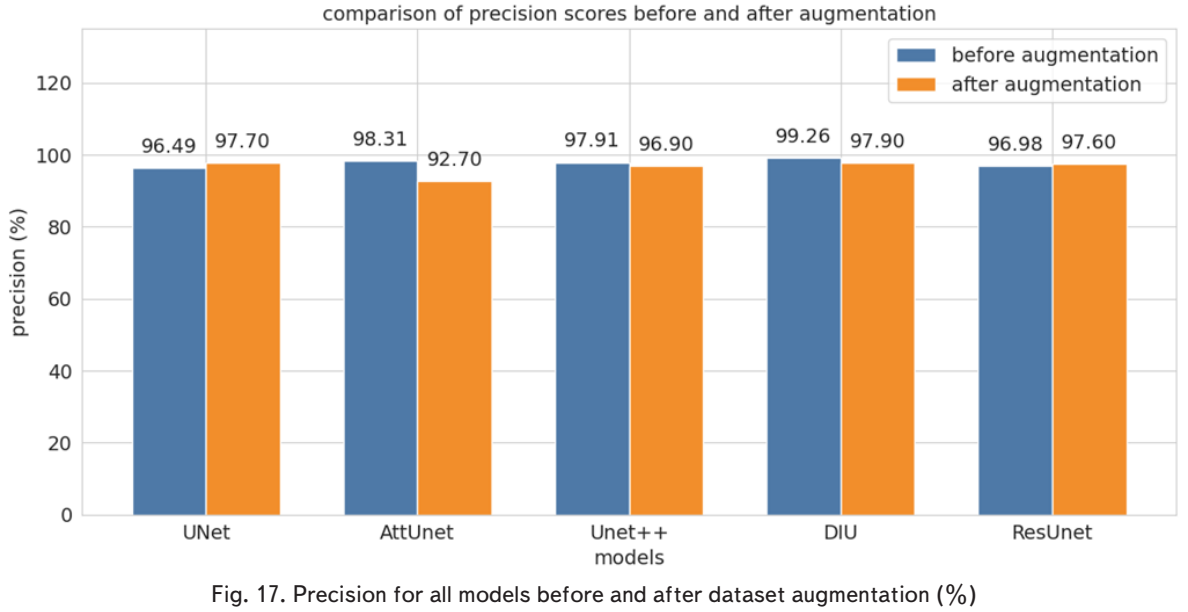

Fig. 17. Precision for all models before and after dataset augmentation (%)

According to the results of the training time on the augmented data shown in Table 4. The learning time of Dense Inception UNet is also higher at 1639069 ms than the other models, also UNet model also has the lowest score 809671 as in the original data.

Table 4

Training time comparison all models on augmented dataset

Model	Training time (ms)
UNet	809671
AttUNet	961035
UNet++	833381
Dense Inception UNet	1639069
ResUNet	840051

Fig. 18 shows the results of breast cancer tumor segment prediction on the original dataset, as it is possible to see that the results of all models match the original mask and the boundaries are clearly delineated.

Bellow Fig. 19 shows the results of all models on the augmented dataset.

Above are derived results of tumor segment prediction on augmented data with added noise and blurring. As it is possible to see from the results all models did well, all borders and fills match the original mask, indicating good performance. Once all the results are displayed, let's go to the results discussion section.

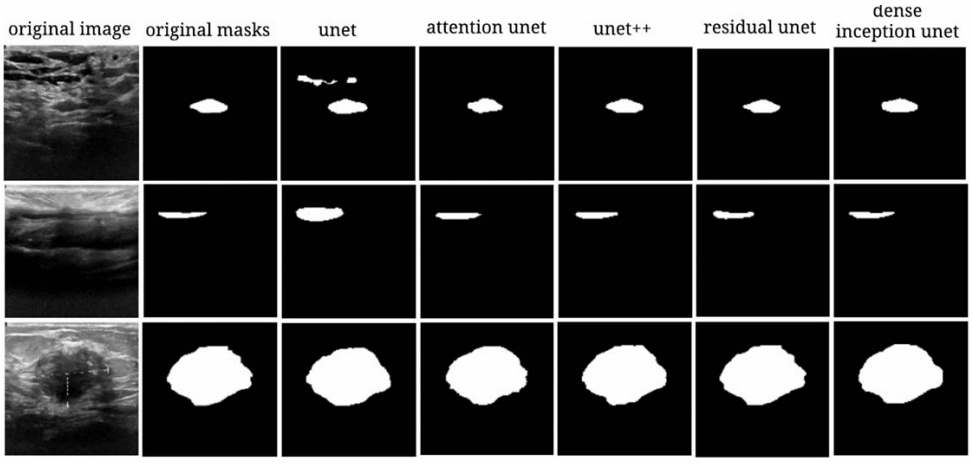

Fig. 18. Cancer prediction results of all models on the original dataset

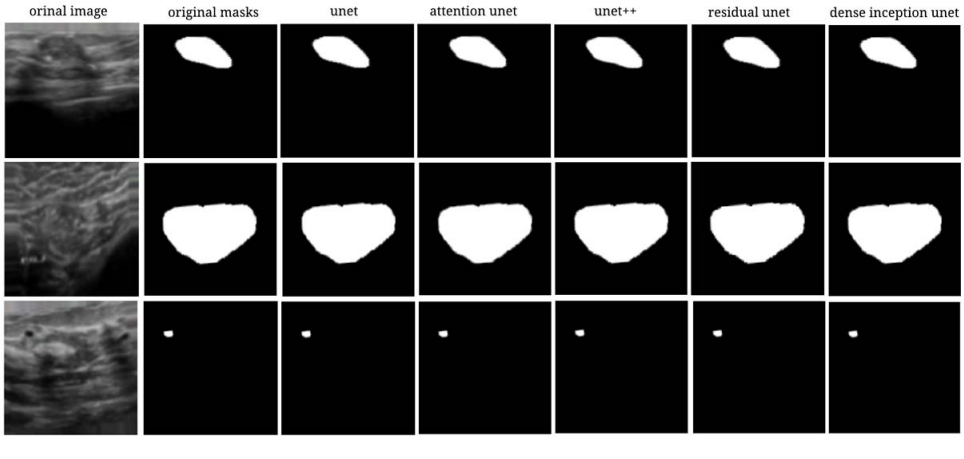

Fig. 19. Breast cancer prediction results by all models on the augmented dataset

6. Discussion of study results of cancer segmentation

The comprehensive evaluation of different breast cancer segmentation models from ultrasound images performed in this study provided valuable insights into their performance. Recent studies in [35, 36] have emphasized the effectiveness of deep learning applications in breast cancer segmentation from ultrasound images compared to classical machine learning methods. These advances play a key role in addressing the problem of early cancer detection and the need for accurate localization of malignancies. However, for better training and improved model accuracy, a large amount of annotated data is required, which may limit the generalization ability of the models in highlighting the region of interest. Synthetic data augmentation methods play an important role in improving the robustness and generalization of deep learning models. In a study [36] experiments were conducted by applying to augmented data, an adaptive histogram equalization method to improve the contrast of images before training the models, yielding improved segmentation accuracy. In this study, it was proposed to introduce speckle noise and fuzziness into the medical ultrasound data, examples of ultrasound images with modifications are shown in Fig. 11. First, the baseline of UNet-based models' performance on raw data was evaluated, with results shown in Table 1. The Dense Inception UNet model showed high segmentation performance on the raw data, obtaining a Dice index of 0.9734, an F1 score of 0.9924 and a minimum loss error of 0.19, which gives the potential of using inception dense modules in this task, as the network can be greatly expanded using the Inception-Res block function and the residual connectivity facilitates the training of the network, in the middle part of the network, the dense connectivity in the Dense-Inception block makes the network deeper and avoids the possible disappearance of the gradient. However, it should be noted that this model requires relatively longer training time than the others. The Residual UNet model also showed excellent segmentation results, with a Dice index of 0.977 and F1 score of 0.976. The UNet model, both before and after augmentation, showed high segmentation results, obtaining a Dice index of 0.9642 and F1 score of 0.9763, with less training time. On the other hand, the Attention UNet model with attention blocks, showed less impressive results, with a Dice index of 0.9681 and F1 score

of 0.9832. Also, Fig.16 and Fig. 17 show the comparison in terms of the Dice coefficient and Precision metrics before and after the injection of noise. UNet showed improved performance on Dice and Precision (1.38 % and 1.21 % increase, respectively). In contrast, UNet's Attention performance decreased significantly (10.81 % on Dice and 5.61 % on Precision). UNet++ showed a slight decrease (1.26 % for Dice and 1.01 % for Precision). Dense Inception UNet showed a slight increase (0.26 % in Dice) and residual UNet showed a decrease (1.39 % in Dice). In [37] a contrast enhancement and speckle noise removal method was applied before training the UNet models, and the training resulted in a Dice coefficient of 0.825. These additions in data processing significantly improved the model training results. In the research work [38], experiments were also conducted with data augmentations, namely using horizontal and vertical cutting, specular reflection and diagonal cutting algorithms resulted in 9.66 % and 12.43 % improvement in segmentation performance of Mask-RCNN and U-Net models. The augmentation method proposed in our study with the addition of speckle noise allowed to identify models that showed stability both before and after the introduction of changes.

In addition, in works [39], a comparative analysis of models based on UNet and semantic convolutional networks for breast ultrasound image tumor segmentation was conducted. The effectiveness of the models was evaluated using metrics such as the Dice coefficient, Jaccard index, accuracy, completeness, specificity, and precision. All models demonstrated Dice coefficients above 75 %. However, the GG-Net and SegResNetVAE models performed better, achieving 82.56 % and 81.90 %, respectively. Furthermore, in the research work [40, 41], additional residue and attention blocks were incorporated into the basic UNet architecture to enhance segmentation. The results showed that the Dice index value was 0.921, slightly lower than that of UNet with dense input, indicating its effectiveness and the promising use of dense and input blocks in this task.

Thus, this study underscores the important role of synthetic data augmentation techniques and UNet-based algorithms. However, limitations include the constraints in data variability, as there are numerous tumor types, and further research is needed on larger annotated datasets. To further improve the study, it is important to consider the possibility of increasing the amount of data and data diversity. Optimizing training time and evaluating the practical applicability of models in real clinical settings also represent important perspectives for future research in medical image processing.

7. Conclusions

1. Comparative evaluation of segmentation methods: in the study, after training the models using Keras library and Tensorflow, different image segmentation approaches were compared to distinguish and highlight healthy breast tissue and cancerous tumors in ultrasound images. The analysis covered various CNN-based models including UNet, Attention UNet, UNet++, Dense Inception UNet and Residual UNet. Among them, the Dense Inception UNet model stands out with a Dice coefficient of 0.9734, F1 score of 0.9924 and minimum loss error of 0.19, indicating its effectiveness in this task.

2. Effect of data augmentation methods: the study explored the impact of data augmentation techniques on breast cancer segmentation accuracy. Expanding the dataset to 2000 samples and introducing noisy (Gaussian noise with a standard deviation of 0.05) and blurring techniques resulted in improved model performance. UNet and Dense Inception UNet models consistently demonstrated strong segmentation results both before and after data augmentation, with UNet achieving Dice coefficients of 0.978 and 0.9786, and DIU reaching 0.976 before augmentation.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship, or otherwise, that could affect the research and its results presented in this paper.

Financing

The study was performed without financial support.

Data availability

Data cannot be made available for reasons disclosed in the data availability statement.

References

1. World Health Organization. URL: https://www.who.int/
2. Dunenova, G. A., Kalmataeva, Z. A., Kaidarova, D. R., Shatkovskaya, O. V., Zhylkaidarova, A. Z., Marchenko, E. A., Glushkova, N. E. (2023). Breast cancer epidemiology in Kazakhstan for the period 2012-2021. Science & Healthcare, 25 (2). doi: https://doi.org/10.34689/SH.2023.25.2.018
3. Igissinov, N., Toguzbayeva, A., Turdaliyeva, B., Igissinova, G., Bilyalova, Z., Akpolatova, G. et al. (2020). Breast Cancer in Megapolises of Kazakhstan: Epidemiological Assessment of Incidence and Mortality. Iranian Journal of Public Health. doi: https://doi.org/10.18502/ijph.v48i7.2948
4. Barba, D., León-Sosa, A., Lugo, P., Suquillo, D., Torres, F., Surre, F. et al. (2021). Breast cancer, screening and diagnostic tools: All you need to know. Critical Reviews in Oncology/Hematology, 157, 103174. doi: https://doi.org/10.1016/j.critrevonc.2020.103174
5. Ginsburg, O., Yip, C., Brooks, A., Cabanes, A., Caleffi, M., Dunstan Yataco, J. A. et al. (2020). Breast cancer early detection: A phased approach to implementation. Cancer, 126 (S10), 2379–2393. doi: https://doi.org/10.1002/cncr.32887
6. Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A. et al. (2021). Deep learning-enabled medical computer vision. Npj Digital Medicine, 4 (1). doi: https://doi.org/10.1038/s41746-020-00376-2

7. Dhahri, H., Al Maghayreh, E., Mahmood, A., Elkilani, W., Faisal Nagi, M. (2019). Automated Breast Cancer Diagnosis Based on Machine Learning Algorithms. Journal of Healthcare Engineering, 2019, 1–11. doi: https://doi.org/10.1155/2019/4253641
8. MI-STA 2022 Conference Proceeding. (2022). 2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA). doi: https://doi.org/10.1109/mi-sta54861.2022.9837515
9. Kuttan, G. O., Elayidom, M. S. (2023). Review on Computer Aided Breast Cancer Detection and Diagnosis using Machine Learning Methods on Mammogram Image. Current Medical Imaging Formerly Current Medical Imaging Reviews, 19 (12). doi: https://doi.org/10.2174/1573405619666230213093639
10. Liu, Z., Wang, S., Dong, D., Wei, J., Fang, C., Zhou, X. et al. (2019). The Applications of Radiomics in Precision Diagnosis and Treatment of Oncology: Opportunities and Challenges. Theranostics, 9 (5), 1303–1322. doi: https://doi.org/10.7150/thno.30309
11. Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, 234–241. doi: https://doi.org/10.1007/978-3-319-24574-4_28
12. Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., Liang, J. (2018). UNet++: A Nested U-Net Architecture for Medical Image Segmentation. Lecture Notes in Computer Science, 3–11. doi: https://doi.org/10.1007/978-3-030-00889-5_1
13. Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K. et al. (2018). Attention u-net: Learning where to look for the pancreas. arXiv. doi: https://doi.org/10.48550/arXiv.1804.03999
14. Zhang, Z., Liu, Q., Wang, Y. (2018). Road Extraction by Deep Residual U-Net. IEEE Geoscience and Remote Sensing Letters, 15 (5), 749–753. doi: https://doi.org/10.1109/lgrs.2018.2802944
15. Targ, S., Almeida, D., Lyman, K. (2016). Resnet in resnet: Generalizing residual architectures. arXiv. doi: https://doi.org/10.48550/arXiv.1603.08029
16. Zhang, Z., Wu, C., Coleman, S., Kerr, D. (2020). DENSE-INception U-net for medical image segmentation. Computer Methods and Programs in Biomedicine, 192, 105395. doi: https://doi.org/10.1016/j.cmpb.2020.105395
17. Zhao, Y., Lai, Z., Shen, L., Kong, H. (2022). Breast Lesions Segmentation using Dual-level UNet (DL-UNet). 2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS). doi: https://doi.org/10.1109/cbms55023.2022.00067
18. Honghan, Z., Liu, D. C., Jingyan, L., Liu, P., Yin, H., Peng, Y. (2021). RMS-SE-UNet: A Segmentation Method for Tumors in Breast Ultrasound Images. 2021 IEEE 6th International Conference on Computer and Communication Systems (ICCCS). doi: https://doi.org/10.1109/icccs52626.2021.9449302
19. Negi, A., Raj, A. N. J., Nersisson, R., Zhuang, Z., Murugappan, M. (2020). RDA-UNET-WGAN: An Accurate Breast Ultrasound Lesion Segmentation Using Wasserstein Generative Adversarial Networks. Arabian Journal for Science and Engineering, 45 (8), 6399–6410. doi: https://doi.org/10.1007/s13369-020-04480-z
20. Chavan, T., Prajapati, K., JV, K. R. (2022). InvUNET: Involuted UNET for Breast Tumor Segmentation from Ultrasound. Lecture Notes in Computer Science, 283–290. doi: https://doi.org/10.1007/978-3-031-09342-5_27
21. Chen, G., Li, L., Dai, Y., Zhang, J., Yap, M. H. (2023). AAU-Net: An Adaptive Attention U-Net for Breast Lesions Segmentation in Ultrasound Images. IEEE Transactions on Medical Imaging, 42 (5), 1289–1300. doi: https://doi.org/10.1109/tmi.2022.3226268
22. Agarap, A. F. (2018). Deep Learning using Rectified Linear Units (ReLU). arXiv. doi: https://doi.org/10.48550/arXiv.1803.08375
23. Liang, X., Wang, X., Lei, Z., Liao, S., Li, S. Z. (2017). Soft-Margin Softmax for Deep Classification. Lecture Notes in Computer Science, 413–421. doi: https://doi.org/10.1007/978-3-319-70096-0_43
24. Guerrero-Pena, F. A., Marrero Fernandez, P. D., Ing Ren, T., Yui, M., Rothenberg, E., Cunha, A. (2018). Multiclass Weighted Loss for Instance Segmentation of Cluttered Cells. 2018 25th IEEE International Conference on Image Processing (ICIP). doi: https://doi.org/10.1109/icip.2018.8451187
25. Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., Rueckert, D. (2019). Attention gated networks: Learning to leverage salient regions in medical images. Medical Image Analysis, 53, 197–207. doi: https://doi.org/10.1016/j.media.2019.01.012
26. Khanna, A., Londhe, N. D., Gupta, S., Semwal, A. (2020). A deep Residual U-Net convolutional neural network for automated lung segmentation in computed tomography images. Biocybernetics and Biomedical Engineering, 40 (3), 1314–1327. doi: https://doi.org/10.1016/j.bbe.2020.07.007
27. He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: https://doi.org/10.1109/cvpr.2016.90
28. Chicco, D., Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics, 21 (1). doi: https://doi.org/10.1186/s12864-019-6413-7
29. Shamir, R. R., Duchin, Y., Kim, J., Sapiro, G., Harel, N. (2018). Continuous Dice Coefficient: a Method for Evaluating Probabilistic Segmentations. doi: https://doi.org/10.1101/306977
30. Derczynski, L. (2016). Complementarity, F-score, and NLP Evaluation. Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). Portorož, 261–266. URL: https://aclanthology.org/L16-1040/
31. Eelbode, T., Bertels, J., Berman, M., Vandermeulen, D., Maes, F., Bisschops, R., Blaschko, M. B. (2020). Optimization for Medical Image Segmentation: Theory and Practice When Evaluating With Dice Score or Jaccard Index. IEEE Transactions on Medical Imaging, 39 (11), 3679–3690. doi: https://doi.org/10.1109/tmi.2020.3002417
32. Cheng, B., Girshick, R., Dollar, P., Berg, A. C., Kirillov, A. (2021). Boundary IoU: Improving Object-Centric Image Segmentation Evaluation. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). doi: https://doi.org/10.1109/cvpr46437.2021.01508

33. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A. (2020). Dataset of breast ultrasound images. Data in Brief, 28, 104863. doi: https://doi.org/10.1016/j.dib.2019.104863
34. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A. (2019). Deep Learning Approaches for Data Augmentation and Classification of Breast Masses using Ultrasound Images. International Journal of Advanced Computer Science and Applications, 10 (5). doi: https://doi.org/10.14569/ijacsa.2019.0100579
35. Akkus, Z., Cai, J., Boonrod, A., Zeinoddini, A., Weston, A. D., Philbrick, K. A., Erickson, B. J. (2019). A Survey of Deep-Learning Applications in Ultrasound: Artificial Intelligence–Powered Ultrasound for Improving Clinical Workflow. Journal of the American College of Radiology, 16 (9), 1318–1328. doi: https://doi.org/10.1016/j.jacr.2019.06.004
36. Ilesanmi, A. E., Chaumrattanakul, U., Makhanov, S. S. (2021). A method for segmentation of tumors in breast ultrasound images using the variant enhanced deep learning. Biocybernetics and Biomedical Engineering, 41 (2), 802–818. doi: https://doi.org/10.1016/j.bbe.2021.05.007
37. Almajalid, R., Shan, J., Du, Y., Zhang, M. (2018). Development of a Deep-Learning-Based Method for Breast Ultrasound Image Segmentation. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). doi: https://doi.org/10.1109/icmla.2018.00179
38. Zhang, C., Bao, N., Sun, H., Li, H., Li, J., Qian, W., Zhou, S. (2022). A Deep Learning Image Data Augmentation Method for Single Tumor Segmentation. Frontiers in Oncology, 12. doi: https://doi.org/10.3389/fonc.2022.782988
39. Ferreira, M. R., Torres, H. R., Oliveira, B., Gomes-Fonseca, J., Morais, P., Novais, P., Vilaca, J. L. (2022). Comparative Analysis of Current Deep Learning Networks for Breast Lesion Segmentation in Ultrasound Images. 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). doi: https://doi.org/10.1109/embc48229.2022.9871091
40. Mukasheva, A., Akanov, Z., Yedilkhan, D. (2021). Research of the Regression Analysis Methods for Predicting the Growth of Patients with Diabetes Mellitus. 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST). doi: https://doi.org/10.1109/sist50301.2021.9465975
41. Zhao, T., Dai, H. (2022). Breast Tumor Ultrasound Image Segmentation Method Based on Improved Residual U-Net Network. Computational Intelligence and Neuroscience, 2022, 1–9. doi: https://doi.org/10.1155/2022/3905998