scientific reports

Check for updates

OPEN

Attention based UNet model for breast cancer segmentation using BUSI dataset

Adel Sulaiman[1,2], Vatsala Anand[3], Sheifali Gupta[3], Adel Rajab[1], Hani Alshahrani[1,2], Mana Saleh Al Reshan[2,4], Asadullah Shaikh[2,4]✉ & Mohammed Hamdi[1,2]

Breast cancer, a prevalent and life-threatening disease, necessitates early detection for the effective intervention and the improved patient health outcomes. This paper focuses on the critical problem of identifying breast cancer using a model called Attention U-Net. The model is utilized on the Breast Ultrasound Image Dataset (BUSI), comprising 780 breast images. The images are categorized into three distinct groups: 437 cases classified as benign, 210 cases classified as malignant, and 133 cases classified as normal. The proposed model leverages the attention-driven U-Net's encoder blocks to capture hierarchical features effectively. The model comprises four decoder blocks which is a pivotal component in the U-Net architecture, responsible for expanding the encoded feature representation obtained from the encoder block and for reconstructing spatial information. Four attention gates are incorporated strategically to enhance feature localization during decoding, showcasing a sophisticated design that facilitates accurate segmentation of breast tumors in ultrasound images. It displays its efficacy in accurately delineating and segregating tumor borders. The experimental findings demonstrate outstanding performance, achieving an overall accuracy of 0.98, precision of 0.97, recall of 0.90, and a dice score of 0.92. It demonstrates its effectiveness in precisely defining and separating tumor boundaries. This research aims to make automated breast cancer segmentation algorithms by emphasizing the importance of early detection in boosting diagnostic capabilities and enabling prompt and targeted medical interventions.

Keywords Breast, Segmentation, Attention-driven U-Net, Breast ultrasound images (BUSI), Intersection over Union (IoU)

Breast cancer is a type of cancer that develops in the cells of the breast. It is the most frequent cancer in women globally and can also strike men. Breast cancer can originate in several regions of the breast, such as the glands that create milk (lobular carcinoma), the ducts that transport milk to the nipple (ductal carcinoma), or other cells in the breast.

Breast cancer is typically detected as a lump in the breast or through screening mammography, which can detect changes in the breast tissue before a lump can be felt. About 2.3 million new cases of breast cancer are expected to be diagnosed worldwide by the year 2020[1,2].

Traditional methods for diagnosing breast cancer depend intensely on imaging procedures such as mammography, ultrasound, and MRI, as well as tissue biopsy for definitive determination. These traditional methods have been instrumental within the early discovery and determination of breast cancer, permitting for convenient treatment and improved results. However, they can be time-consuming, expensive, and may now and then lead to false-positive or false-negative comes about, highlighting the require for more progressed and exact diagnostic approaches[3,4].

Semantic segmentation, a computer vision technique, holds critical promise for progressing the determination of breast cancer. By precisely depicting and classifying the diverse locales inside breast tissue, such as tumors, ducts, and stroma, semantic segmentation can help radiologists and pathologists in recognizing and characterizing lesions more successfully. This innovation can analyze mammograms, MRI scans, or histopathology images,

[1]Department of Computer Science, College of Computer Science and Information Systems, Najran University, Najran 61441, Saudi Arabia. [2]Emerging Technologies Research Lab (ETRL), College of Computer Science and Information Systems, Najran University, Najran 61441, Saudi Arabia. [3]Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab 140401, India. [4]Department of Information System, College of Computer Science and Information Systems, Najran University, Najran 61441, Saudi Arabia. ✉email: asshaikh@nu.edu.sa

giving detailed data around the measure, shape, and area of anomalies. This level of detail can lead to prior and more precise detection of breast cancer, possibly empowering mediations at prior stages when treatment is more compelling. Here, a semantic segmentation based attention U-Net model is proposed for breast cancer detection that can reduce the subjectivity and variability associated with manual interpretation, enhancing the consistency and reliability of diagnoses. As a result, proposed model has the potential to revolutionize breast cancer diagnosis by offering a more precise, efficient, and standardized approach to analyze medical images[5–7].

The remaining research is shown as: The literature review is shown in section "Literature review", the input dataset is presented in section "Input dataset", the proposed attention-driven u-net model is given in section "Proposed attention-driven U-Net model", the findings and discussion are presented in section "Results and discussion", and the conclusion is shown in section "Conclusions and future work".

Literature review

The existing literature approaches for breast segmentation are studied here. Chen et al.[8] have presented a hybrid adaptive attention module (HAAM) based U-Net model for the segmentation of breast nodules from ultrasound images. They had worked on two datasets namely BUSI and dataset B with a total of 780 and 163 ultrasound images of breast cancer respectively and obtained the values of precision as 79.61 and 78.83 respectively. The HAAM module in[8] introduces a more complex attention mechanism, which may require more data or tuning to generalize well to the segmentation task compared to the simpler attention gates. Punn et al.[9] presented an attention-guided U-Net model for breast cancer segmentation. They worked using the BUSI dataset and obtained the values of accuracy, precision, and recall as 0.970, 0.940, and 0.891 respectively. It introduces residual inception depth-wise separable convolution and hybrid pooling layers. Its complex architecture can make it harder to interpret and understand how the model is making its predictions. Li et al.[10] presented U2-MNet for the segmentation of breast cancer. They worked using the BUSI dataset and obtained the values of dice, IOU and Precision as 0.8276, 0.7317 and 0.8624 respectively. Zhu et al.[11] introduced the SU-Next model as a method for segmenting breast tumors. They utilized the BUSI dataset and derived the values of IoU and dice as 0.6426 and 0.7740, respectively. The U-Next network replaces the last two convolutional blocks of the encoding stage of the U-Net network with the last two convolutional blocks of decoding with a multilayer perceptron structure. While the modifications to the U-Net architecture may improve segmentation speed, especially by replacing convolutional blocks with a multilayer perceptron (MLP) structure, there could be a risk of losing some level of segmentation accuracy or quality.

In 2022, Podda et al.[12] introduced a model that utilized photos from the BUSI dataset to do breast cancer segmentation. The dice score they earned was 0.826. Huang et al.[13] demonstrated the segmentation of breast ultrasonography lesions utilizing BUS images. They acquired the Dice score values as 0.854 and IOU as 0.919. Chen et al.[14] introduced a cascaded convolutional neural network to segment breast cancer. They utilized the BUSI dataset for their investigation. Tang et al.[15] introduced a feature pyramid nonlocal network to segment breast tumors. The dice scores achieved on two distinct datasets were 0.847 and 0.87. Xue et al.[16] demonstrated the process of segmenting breast lesions from ultrasound images. By utilizing the BUSI dataset, they derived the dice score, recall, and accuracy values as 0.954, 0.957, and 0.951, respectively. Mishra et al.[17] introduced a machine-learning technique for detecting malignancy in breast tumors. By utilizing the BUSI dataset, they attained accuracy and F1-score values of 0.974 and 0.94, respectively. Xing et al.[18] introduced three pre-trained models specifically designed for the segmentation of breast tumors. The accuracy value they obtained was 0.843. Byra et al.[19] introduced a technique for scaling deep representations to perform breast tumor segmentation and achieved an accuracy of 0.915. The surveyed literature primarily focuses on U-Net variations and convolutional neural networks. Exploring a broader range of architectures and comparing their effectiveness could contribute to advancements in breast segmentation techniques. This study makes several noteworthy contributions to the field, which are given below:

1. This research introduces a novel Attention-driven U-Net model for breast cancer segmentation, featuring meticulously crafted architecture. The proposed model leverages the attention-driven U-Net's encoder blocks to capture hierarchical features effectively.
2. The model comprises four encoder blocks, progressively reducing spatial dimensions and increasing channel depth that ensures the extraction of intricate details crucial for breast cancer segmentation.
3. The model comprises four decoder blocks which is a pivotal component in the U-Net architecture, responsible for expanding the encoded feature representation obtained from the encoder block and for reconstructing spatial information.
4. Four attention gates are incorporated strategically to enhance feature localization during decoding, showcasing a sophisticated design that facilitates accurate segmentation of breast tumors in ultrasound images. Attention gates facilitate improved feature localization by assigning attention weights to different regions of the feature maps.

Input dataset

The Breast Ultrasound Image dataset (BUSI) is a collection of ultrasound images of the breast, annotated by experts to indicate regions of interest such as tumors, cysts, and normal tissue. The annotations provide ground truth labels that can be used to train and evaluate segmentation algorithms. Researchers use the BUSI dataset to develop and validate automated methods for segmenting breast lesions, for early detection of breast cancer.

This study uses a collection of Breast Ultrasound Images (BUSI) to find and outline areas of breast cancer. The baseline data contains breast ultrasound images that were collected from an average of 600 female patients. The collection has 780 images that have been sorted into different groups. Out of these, 210 images are of malignant breast cancer, 437 depict cases of benign breast cancer, and 133 depict normal breast cancer images. Every

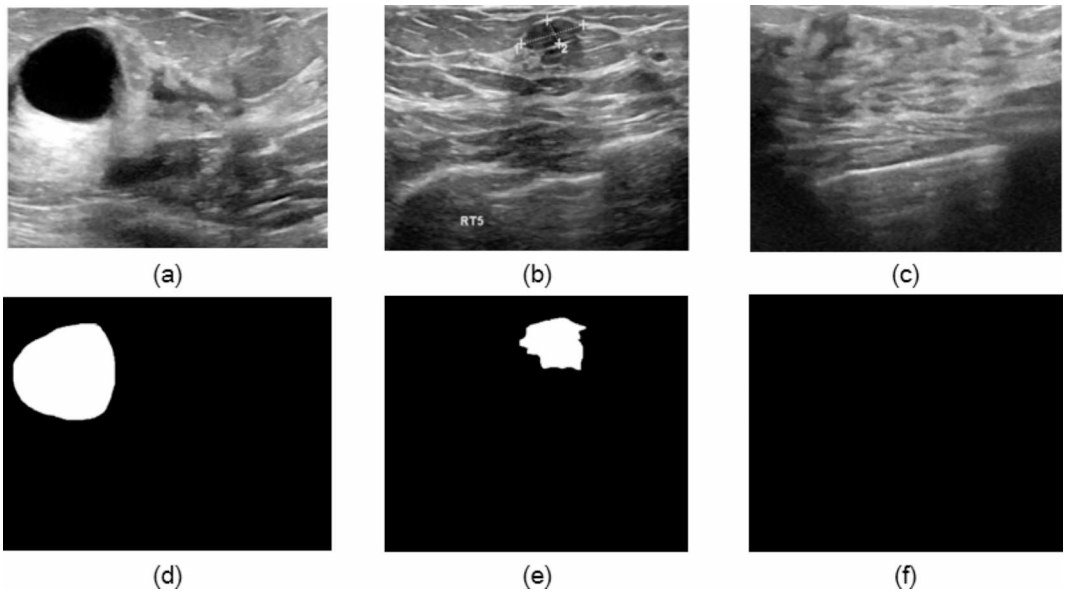

Fig. 1. Dataset samples (**a**) benign, (**b**) malignant, (**c**) normal image without tumor, (**d**) benign mask, (**e**) malignant mask, (**f**) normal mask without any tumor.

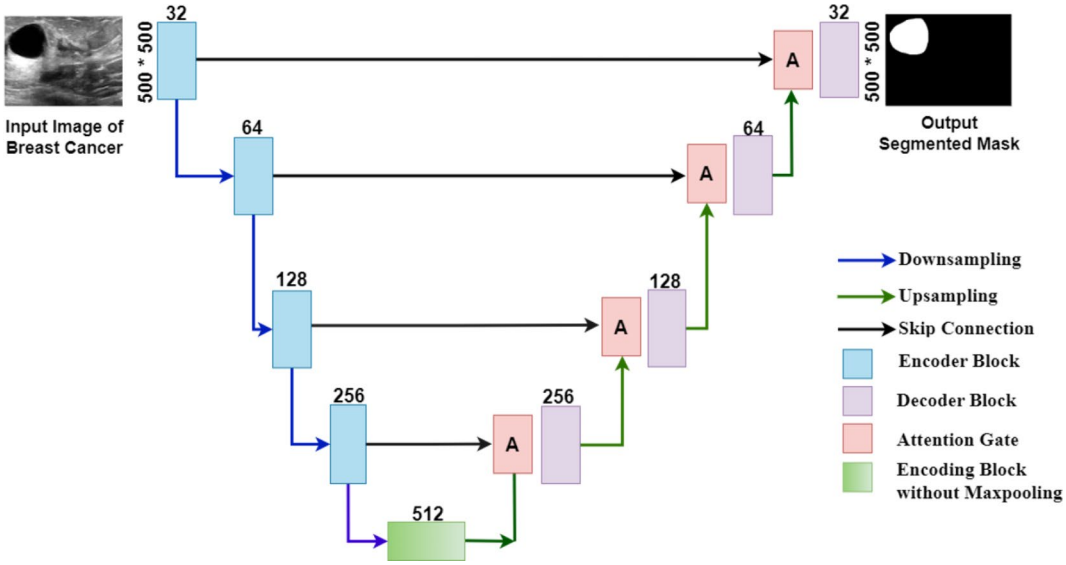

Fig. 2. Attention-driven U-net model.

single image in the collection is saved in the PNG format and has an average size of 500 pixels * 500 pixels. Data Augmentation technique is used to balance the dataset by creating modified versions of images in the dataset. In this study, augmentation is performed to increase the number of images to 1500, with 500 images for each class (malignant, benign, and normal breast images), by creating variations of the original images. Figure 1a depicts the benign image sample, 1b depicts the malignant image sample, 1c shows normal image without tumor, 1d-f shows the masks of benign, malignant and normal images respectively.

Proposed attention-driven U-Net model

Figure 2 depicts the proposed Attention-driven U-Net architecture that is utilized for breast cancer segmentation. The "attention-driven" model refers to a model architecture that includes mechanisms to focus on specific parts of the input data, giving them more "attention" during processing. This attention mechanism allows the model to selectively emphasize relevant features while suppressing irrelevant or distracting information.

The proposed U-Net model with attention-driven mechanism has symmetrical architecture with an encoder pathway and a decoder pathway. It has four encoders and four decoders with an attention gate mechanism. The Attention Gate layers are inserted between the encoding and decoding stages of the U-Net to enhance the

model's ability to focus on relevant features while ignoring irrelevant ones. Encoder blocks downsample the input image to extract high-level features. Each encoder block consists of convolutional layers followed by a pooling operation. Attention gates are applied before each decoder block to selectively highlight important features from the encoding stage. These attention gates help the decoder focus on relevant features for better segmentation. Decoder blocks upsample the features and concatenate them with features from the corresponding encoder block, followed by convolutional layers.

The encoder pathway contracts the input image, capturing high-level features. The decoder pathway expands the features back to the original input image size. In Fig. 2, the encoder pathway is on the left side of the diagram and the decoder pathway is on the right side. The encoder pathway consists of several convolutional blocks that contract the input image. Each convolutional block contains a convolutional layer, a batch normalization layer, and a ReLU activation layer. The decoder pathway consists of several upsampling blocks that expand the features back to the original input image size.

Attention gates and skip connections are also used in the proposed architecture. Skip connections are connections that skip over different layers in the model. They allow the decoder pathway to access information from the earlier stages of the encoder pathway. By using attention gates, CNNs can concentrate on the most informative parts of the data, leading to potentially better performance in tasks like image segmentation or object detection.

Figure 3 shows an overview of the model which explains about each used layer, its output shape and number of parameters for each layer respectively. The detailed description of encoder block, attention gate and decoder block is given in upcoming sections.

Encoder block

The Encoder Block is the most important part of the U-Net design used to detect breast cancer. This part helps to extract and arrange details from the ultrasound images to make them easier to use. The author has implemented the proposed model with four encoders and attention gates as compared to Oktay et al.[2] who has implemented the attention UNet with three encoders to segment pancreas. It causes an increase in the processing time due to extra encoder but the proposed model with four encoders represents the best trade-off between performance and processing time. The purpose of the Encoder Block, which consists of two convolutional layers, Rectified Linear unit (ReLU) activation, and dropout regularization after the first convolution layer, is to identify and amplify important patterns in the data, as illustrated in Fig. 4. The feature extraction process is started by the 1st convolutional layer, which has a 3×3 filter size and uses the 'same' padding approach, as shown in Eq. (1).

$$c1\_output = ReLU(Conv2D(X,\ filters,\ (3,3),1,padding = 'same')) \tag{1}$$

The input feature map is labelled as X, and the output of the first convolution layer is called c1_output. The dropout layer includes a stochastic element to stop overfitting. The dropout output is presented in Eq. (2).

$$dropout\_output = Dropout\ (c1\_output,\ dropout\_rate) \tag{2}$$

Equation (3) illustrates the second convolutional layer, which enhances the learning of complex spatial representations by further refining these characteristics. This convolution layer has a 3×3 kernel size and uses the 'same' padding approach.

$$c2\_output = ReLU(Conv2D(dropout\_output,\ filters,(3,3),1,padding = 'same')) \tag{3}$$

where dropout_output is the feature map obtained from the dropout layer and c2_output is the output of the second convolution layer.

Moreover, the optional max-pooling layer of the Encoder Block facilitates spatial downsampling and feature reduction, hence aiding in the efficient collection of hierarchical information. The pooling process illustrated in Eq. (4) helps to maintain important contextual information, making sure that the encoder successfully gathers and saves pertinent data for later stages of decoding.

$$final\_output = MaxPool2D\,(c2\_output) \tag{4}$$

The Encoder Block's flexibility in accommodating different architectural specifications is demonstrated by its capacity to either incorporate or eliminate pooling. Overall, by making it simpler for the model to accurately encode and represent crucial information from breast ultrasound images.

Attention gate block

The Attention Gate block, which is designed expressly to enhance the data flow between the encoder and decoder sections, is a crucial component of the proposed U-Net model, as illustrated in Fig. 5. The Attention Gate gives the U-Net model the ability to adaptively respond to pertinent information. Although segmentation efficiency may be enhanced by adding more attention gate blocks, it is important to weigh the trade-offs in terms of computational expense and overfitting risk. Here, four attention gates are used corresponding to four encoders and decoders. The Attention Gate carefully integrates information from both routes, which represent the encoder and decoder layers' two input branches, to enhance the segmentation process by helping the model concentrate on pertinent regions. A sequence of element-wise and convolutional procedures is used to implement the attention mechanism.

The output from (n-1)th decoder is fed through convolutional route, as indicated by Eq. (5), during the forward pass.

Layer (type)	Output Shape	Param #	Connected to
input_2 (InputLayer)	[(None, 256, 256, 3)	0	
Encoder1 (EncoderBlock)	((None, 128, 128, 32	10144	input_2[0][0]
Encoder2 (EncoderBlock)	((None, 64, 64, 64),	55424	Encoder1[0][0]
Encoder3 (EncoderBlock)	((None, 32, 32, 128)	221440	Encoder2[0][0]
Encoder4 (EncoderBlock)	((None, 16, 16, 256)	885248	Encoder3[0][0]
Encoding (EncoderBlock)	(None, 16, 16, 512)	3539968	Encoder4[0][0]
Attention1 (AttentionGate)	(None, 32, 32, 256)	1771265	Encoding[0][0] Encoder4[0][1]
Decoder1 (DecoderBlock)	(None, 32, 32, 256)	2359808	Encoding[0][0] Attention1[0][0]
Attention2 (AttentionGate)	(None, 64, 64, 128)	443265	Decoder1[0][0] Encoder3[0][1]
Decoder2 (DecoderBlock)	(None, 64, 64, 128)	590080	Decoder1[0][0] Attention2[0][0]
Attention3 (AttentionGate)	(None, 128, 128, 64)	111041	Decoder2[0][0] Encoder2[0][1]
Decoder3 (DecoderBlock)	(None, 128, 128, 64)	147584	Decoder2[0][0] Attention3[0][0]
Attention4 (AttentionGate)	(None, 256, 256, 32)	27873	Decoder3[0][0] Encoder1[0][1]
Decoder4 (DecoderBlock)	(None, 256, 256, 32)	36928	Decoder3[0][0] Attention4[0][0]
conv2d_61 (Conv2D)	(None, 256, 256, 1)	33	Decoder4[0][0]

Total params: 10,200,101
Trainable params: 10,199,141
Non-trainable params: 960

Fig. 3. Model summary of attention-driven U-Net model.

$$x = ReLU(Conv2D(X, filters, 3 \times 3, padding = 'same')) \tag{5}$$

where X is the output from (n-1)th decoder and x is the output of 1st convolution layer.

To match the decoder features' spatial dimensions, the encoder features go through the downsampling process using the second convolution layer described in Eq. (6).

$$skip = ReLU(Conv2D(E, filters, 3 \times 3, 2, padding = 'same')) \tag{6}$$

where E is the encoder output from the skip connection. It is downsampled using a convolution layer enabled with a pooling operation creating output '*skip*'. A fused representation is then created by adding these two sets of features together element-wise as indicated in Eq. (7).

$$z = x + skip \tag{7}$$

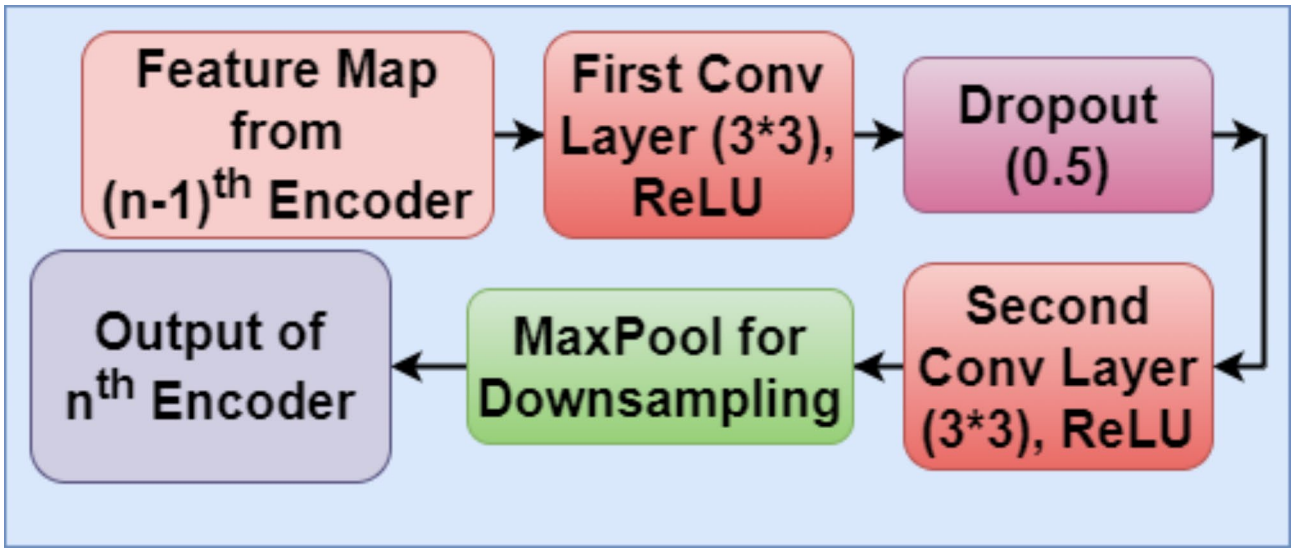

Fig. 4. Block diagram of encoder block.

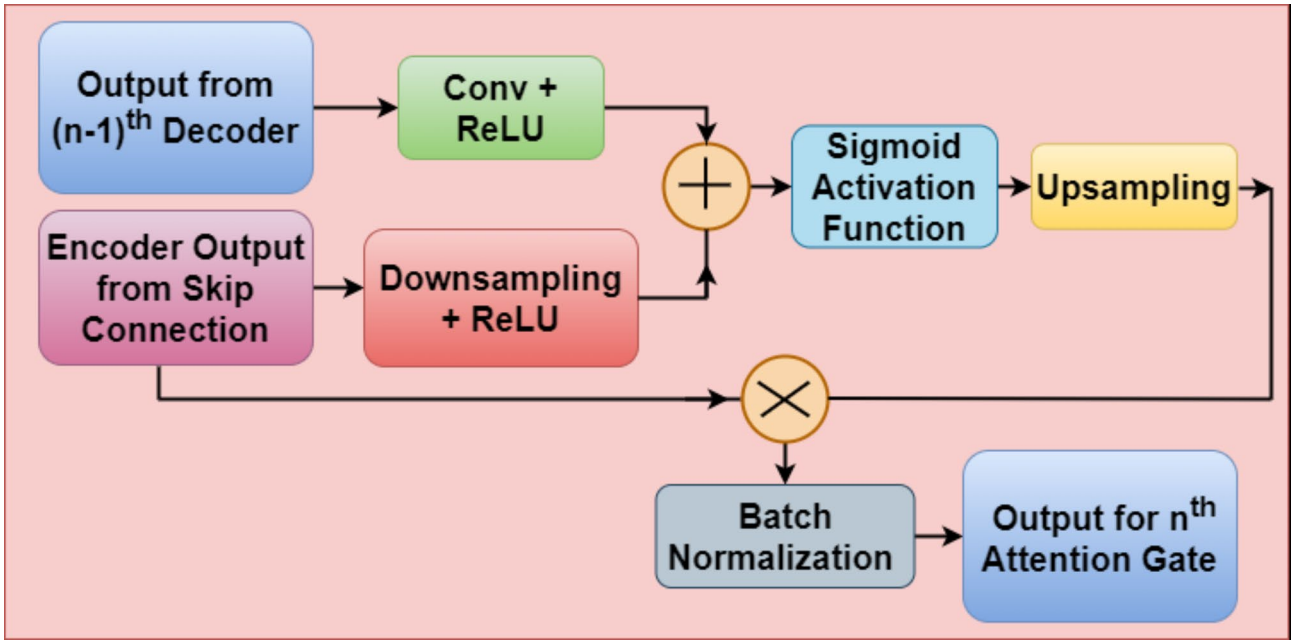

Fig. 5. Attention gate block.

Using a convolutional layer with sigmoid activation as demonstrated in Eq. (8), these fusion representations 'z' are further processed to learn attention weights through convolution layers, resulting in attention maps that highlight important regions.

$$y = sigmoid(Conv2D(z, 1 \times 1, padding = \prime same\prime)) \tag{8}$$

For matching the dimensions of 'y' with original spatial dimensions, upsampling operation is used presented in Eq. (9),

$$f = Upsample\,(y) \tag{9}$$

The upsampled attention maps 'f' are multiplied element-wise with the encoder output from the skip connection 'E' as shown in Eq. (10).

$$w = f * E \tag{10}$$

Through the selective amplification of pertinent traits, the model can concentrate on crucial regions for segmentation. The attention mechanism's stability and efficiency are further improved by the addition of batch normalization, as demonstrated in Eq. (11).

$$final\_output = BatchNormalization(w) \tag{11}$$

where the '*final_output*' is the output from the attention gate block.

The U-Net model can adaptively attend to prominent features at various scales by integrating Attention Gate blocks, which leads to more accurate and contextually informed segmentation findings in breast ultrasound images.

Decoder block

An essential part of the U-Net architecture, the Decoder Block is in charge of enlarging an encoded feature representation that is acquired from the encoder block. As seen in Fig. 6, this block is put in place to the outcomes of the encoding and attention gate blocks.

The Decoder Block uses the UpSampling2D layer to upsample its input at the start of the forward pass as shown in Eq. (12).

$$p = Upsampling2D(d) \tag{12}$$

where d is the output from the $(n\text{-}1)^{th}$ decoder.

Next, the skip connection from the matching encoder block is concatenated with the upsampled feature map as shown in Eq. (13).

$$q = concatenate([p, a]) \tag{13}$$

where a is the output from the $(n\text{-}1)^{th}$ attention gate.

An essential component of the U-Net architecture is this concatenation, which allows the model to combine contextual knowledge gained during the decoding phase with high-resolution features from the encoding step. The encoder block then processes the concatenated feature map, using convolutional techniques to improve and enhance the feature representation as shown in Eq. (14).

$$r = EncoderBlock(q) \tag{14}$$

where 'r' is the final output of the encoder block as discussed in section "Encoder block".

The output demonstrates the decoder block's role in enhancing segmentation details and reconstructing spatial information[21,22].

The attention gate and encoding block outputs are combined by the Decoder Block to guarantee that the model can highlight pertinent features only and produce correct, context-aware segmentation results in the end.

Results and discussion

The breast cancer segmentation using Attention-Driven U-Net is performed in this research and fascinating results are achieved, evidenced by a well-balanced confusion matrix and classification report. The proposed model has been implemented using Adam optimizer with 32 batch size and 15 epochs. The authors had worked using the NVIDIA Tesla GPU and the training time of the model was 1 h 47 min. The splitting ratio of 80:20 is used for training and testing of dataset.

Architecture ablation analysis

An ablation study involves systematically removing or modifying components of a model architecture to understand their contributions to performance. Oktay et al.[2] suggested Attention Unet model with three

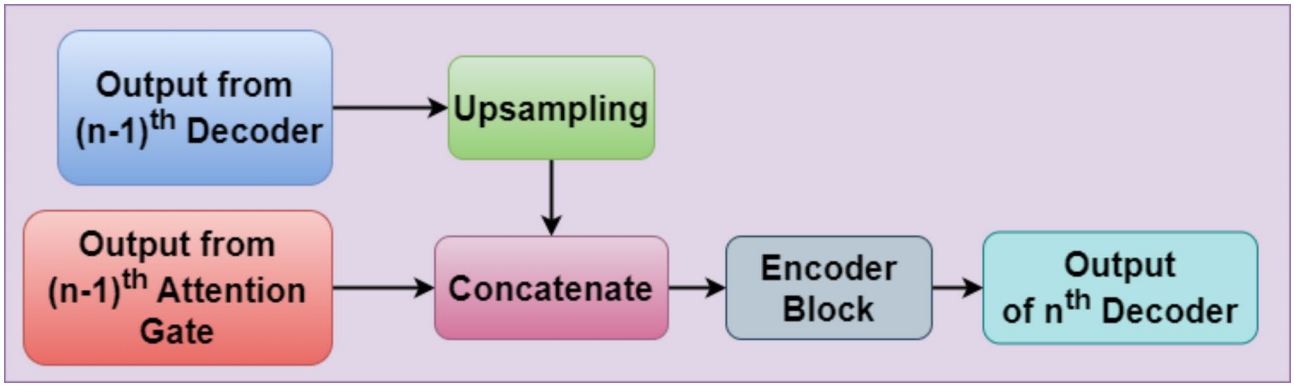

Fig. 6. Decoder block.

Attention-driven U-net with *n* number of encoders where '*n*' is	Precision	Recall/sensitivity	Specificity	Dice score	Jaccard index	Accuracy	Processing time (s)
3	0.84	0.93	0.96	0.77	0.63	0.96	5044
4	**0.96**	**0.89**	**0.99**	**0.85**	**0.73**	**0.98**	**5128**
5	0.93	0.70	0.99	0.55	0.38	0.95	5739

Table 1. Architecture ablation analysis.

	Precision	Recall/sensitivity	Specificity	Dice score	Jaccard index	Accuracy
U-net with Attention	**0.96**	**0.89**	**0.99**	**0.85**	**0.73**	**0.98**
U-net without Attention	0.92	0.87	0.95	0.82	0.70	0.95

Table 2. Comparison of U-Net having four encoders with and without attention.

Kernel size	Precision	Recall/sensitivity	Specificity	Dice score	Jaccard index	Accuracy
3*3	0.96	0.89	0.99	0.85	0.73	0.98
5*5	0.88	0.89	0.98	0.79	0.65	0.97
7*7	0.90	0.77	0.99	0.66	0.50	0.96

Table 3. Parameter ablation analysis of attention-driven U-Net model.

encoders for pancreas segmentation. We implemented the same model on BUSI dataset to segment Breast cancer on ultrasound images and got the accuracy of 96%. U-Net architectures rely on a hierarchy of features extracted at multiple scales. With only three encoders, the depth of the network is limited, potentially reducing the richness of the learned features. Shallow architectures may not capture complex patterns and structures in the data as effectively as deeper architectures. Hence, we implemented the same model by changing the number of Attention gates and encoders from three to four and five because the number of encoders affects the depth and complexity of the features extracted by the model. Optimizing this parameter can help in capturing relevant features for breast cancer segmentation, which can lead to better segmentation results. Hence in this section, the number of encoders are varied to obtain the architecture ablation analysis and the results are depicted in Table 1.

From Table 1, it is analysed that with four encoders, the model showcases best results with precision of 0.96 and overall accuracy of 0.98. Whereas the value of specificity reaches 0.99, followed by recall as 0.89. The Dice Score as 0.85 and Jaccard Index as 0.73 indicates that the model is performing best with four encoders. Moving to five encoders, there is a notable drop in recall at 0.70, whereas the specificity remains the same as 0.99. The Dice Score as 0.55 and Jaccard Index as 0.38 indicate a decrease in segmentation accuracy from 0.98 to 0.95. As the number of encoders increases from 3 to 5, the processing time also increases. Specifically, it goes from 5044 s with 3 encoders to 5739 s with 5 encoders. This indicates that the model becomes slower as more encoders are added. It can be concluded from Table 1 that Attention-driven U-Net model with four encoders represents the best trade-off between performance and processing time, whereas configuration with five encoders suffers from reduced performance metrics and the longest processing time. A well-optimized number of encoders can help the model generalize better to unseen data. It can make the model more robust and capable of segmenting breast cancer in a variety of input images.

Here, the attention mechanism serves as a crucial component, enhancing the U-net's ability to focus on informative regions and improve segmentation accuracy. The study underscores the significance of attention mechanisms in the context of breast cancer segmentation, highlighting their positive impact on precision, recall, and overall segmentation quality within the U-net architecture. Table 2 explains the significance of attention mechanism for breast cancer segmentation to analyse its performance in terms of precision, accuracy and recall. It compares the U-net with attention mechanism that performs well as compared to the U-net without attention. From the Table 2 it is analysed that the attention-driven model performs high with the value of accuracy as 0.98, precision as 0.96, recall as 0.89 and specificity as 0.99. Whereas, the value of Dice Score is 0.84 and the Jaccard Index is 0.73. These results show that the model with attention mechanism helps to capture important details and improve overall performance. On the other hand, the U-net without attention mechanism has a significant fall in precision, which is 0.92, and sensitivity, which is 0.87. A decrease in segmentation quality and overall performance is also reflected by the Dice Score of 0.82, the Jaccard Index of 0.70, and the accuracy of 0.95. It is concluded from the Table 2 that U-Net model with attention mechanism has outperformed as compared to U-Net model without attention. Attention mechanisms allow the model to focus on relevant features while suppressing irrelevant ones. This can improve the quality of segmentation masks by ensuring that the network pays more attention to important regions in the input image. Overall, implementing an attention mechanism in a U-Net architecture can improve the model's performance, sensitivity to details, generalization ability, computational efficiency, and interpretability, making it a valuable addition.

Parameter ablation analysis

In the parameter ablation study, Attention-driven U-net is analysed with different filter sizes (3*3, 5*5, and 7*7) in the attention block which shows notable trends in the performance metrics. In this case, varying the kernel size allows an exploration of how the receptive field size affects the model's ability to capture features at different scales. The findings can guide the selection of optimal architectural configurations for specific tasks, balancing trade-offs between precision, recall, and overall segmentation performance.

From the Table 3, it is analysed that the 3*3 kernel size demonstrates the highest precision 0.96 and specificity 0.99, indicating a superior ability to accurately classify true positive instances and true negative instances. However, this comes at a slight cost to recall 0.89, suggesting a potential trade-off between precision and recall. On the other hand, the 5*5 kernel size exhibits balanced performance across precision, recall, specificity, and accuracy. While slightly lower in precision compared to the 3*3 kernel, it maintains a reasonable balance, resulting in a good overall performance. The 7*7 kernel size shows a decline in several metrics, particularly in recall 0.77, dice score 0.66, and Jaccard index 0.50. This may imply that the larger kernel size introduces challenges in capturing fine details, leading to decreased sensitivity. It is concluded from the Table 3 that Attention-driven U-Net model performs best for the kernel size of 3*3 in attention block.

Proposed model accuracy and loss analysis

From the previous sections, it is observed that the proposed model performs best with Attention-driven mechanism having four encoders with 3*3 kernel size in Attention block. The proposed model's accuracy and loss curves are displayed in Fig. 7. From the Fig. 7a it is analysed that around epoch 8, training accuracy peaks at 95%, then slightly varies before stabilizing. By epoch 12, validation accuracy increases gradually to almost 93%, indicating the model's strong capacity to generalize to new data. This shows that the model is developing applicable features that can be applied to fresh data, rather than simply memorizing the training set.

The attention-driven U-Net model appears to be doing well overall on the job of breast cancer segmentation. These findings hold promise for the model's possible applicability in clinical settings.

According to the model loss curve in Fig. 7b, training loss decreased steadily from epoch 0 to epoch 6, then gradually until epoch 12. This shows that the model is improving its ability to identify breast cancer lesions in ultrasound images and efficiently learn from the training data. Throughout the training phase, the validation loss likewise drops, although it still stays somewhat larger than the training loss. In general, the model loss curve indicates that the Attention-driven U-Net model is a viable method for ultrasound image segmentation of breast cancer.

Confusion matrix parameters analysis

The proposed model uses the U-Net's attention-driven encoder blocks to effectively capture different features. Table 4 shows the confusion matrix parameters analysis that includes precision, accuracy, F1-score and recall results.

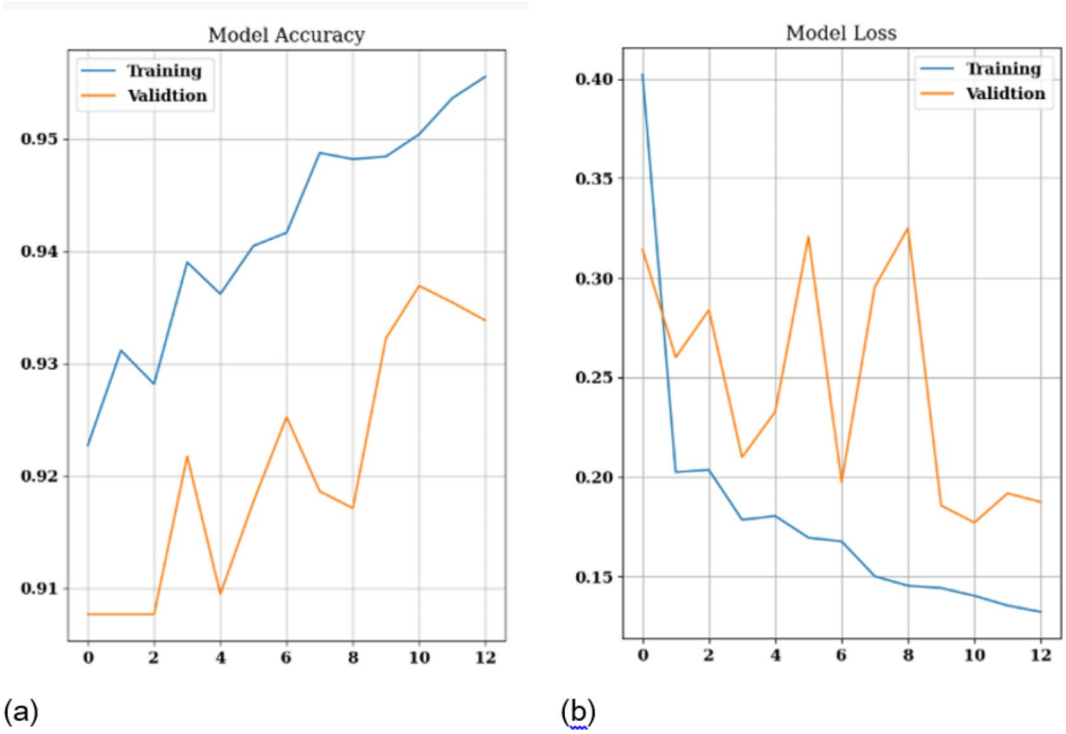

Fig. 7. Training and validation accuracy and loss (**a**) model accuracy, (**b**) model loss.

	Precision	Recall	F1-score	Accuracy
False	0.98	0.99	0.99	0.98
True	0.95	0.80	0.85	
Macro Average	0.97	0.90	0.92	
Weighted Average	0.98	0.98	0.98	

Table 4. Confusion matrix parameters analysis.

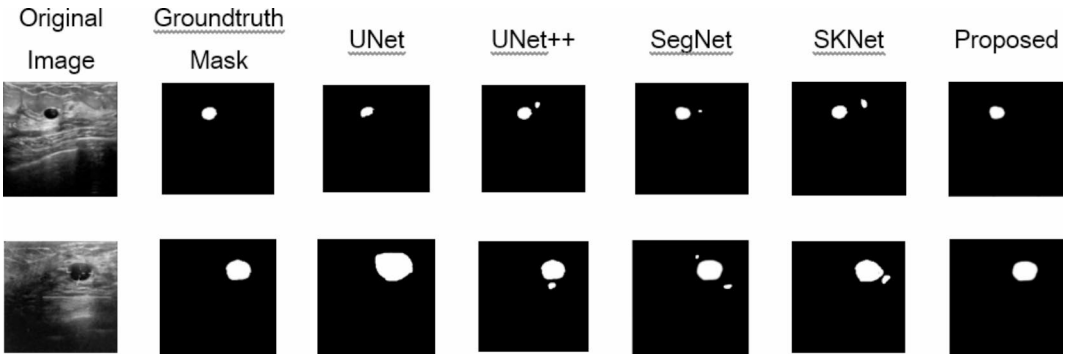

Fig. 8. Segmentation results on different techniques.

The macro average is obtained as 0.97, followed by recall as 0.90 and f1-score as 0.92. The overall accuracy value that is achieved is 0.98. Overall, the confusion matrix and classification report suggest that the Attention-driven U-Net model has the potential to be a valuable tool for breast cancer segmentation in ultrasound images. Its ability to accurately distinguish between healthy and cancerous tissue, coupled with balanced precision and recall, makes it a promising approach for clinical applications.

Qualitative comparison of proposed model with other methods

The performance of proposed model is further compared with other techniques UNet, UNet++, SegNet and SKNet. These all techniques are implemented on BUSI dataset having 780 images. Figure 8 displays the visual segmentation outcomes of various segmentation techniques on the BUSI dataset. In comparison to the segmentation outcomes of existing techniques, the proposed approach not only successfully mitigates the disturbance caused by tumor size, surrounding tissue, and cascade effects, but also attains outcomes for segmentation that are more closely aligned with the ground-truth masks. The proposed study demonstrates superior segmentation outcomes in breast cancer, with less instances of missing and incorrect detections, as indicated by the comprehensive evaluation results and visual effects.

Visualization of proposed model with GradCAM (gradient-weighted class activation mapping)

Grad-CAM is a popular technique for visualizing and interpreting Convolutional Neural Networks (CNNs) by highlighting important regions in the input image that influence the model's prediction. Figure 9 shows the visualization results of the proposed model with GradCAM. Grad-CAM uses the gradients flowing into the final convolutional layer to produce a coarse localization map highlighting the important regions in the image for predicting the concept. Here red shows how much deeper the proposed model has focussed on the affected region.

Comparison with state-of-art

The Table 5 presents a comparative analysis of various state-of-the-art models used in the field, including Adaptive Attention U-Net, residual cross-spatial attention-guided inception U-Net model, U2-MNet, SU-Next, CNN, Selective Kernel U-Net, Soft and hard attention mechanism, and Sub-region Pooling Network. Each technique is evaluated based on its performance using metrics such as Dice Score, Precision, Recall, Specificity, Jaccard Index, and Accuracy. In terms of performance, the proposed model achieves a Dice Score of 0.85, which is slightly lower than[9] model's Dice Score of 0.91. However, the proposed model demonstrates higher Precision as 0.97 vs. 0.94 and similar Recall as 0.90 vs. 0.89 compared to[9] model. This suggests that the proposed model is more precise in identifying positive instances while maintaining a comparable ability to find all positive instances.

In terms of precision and recall, the proposed attention-driven U-net outperforms the Adaptive Attention U-Net[8] in every metric when compared to alternative methods. Furthermore, with significantly higher Dice scores, it outperforms U2-MNet[10], SU-Next[11], CNN[16], and the Selective Kernel U-Net[23].

The proposed attention-driven U-net methodology achieves cutting-edge outcomes in breast cancer segmentation, proving its effectiveness. The research presents the attention-driven U-net as a strong contender

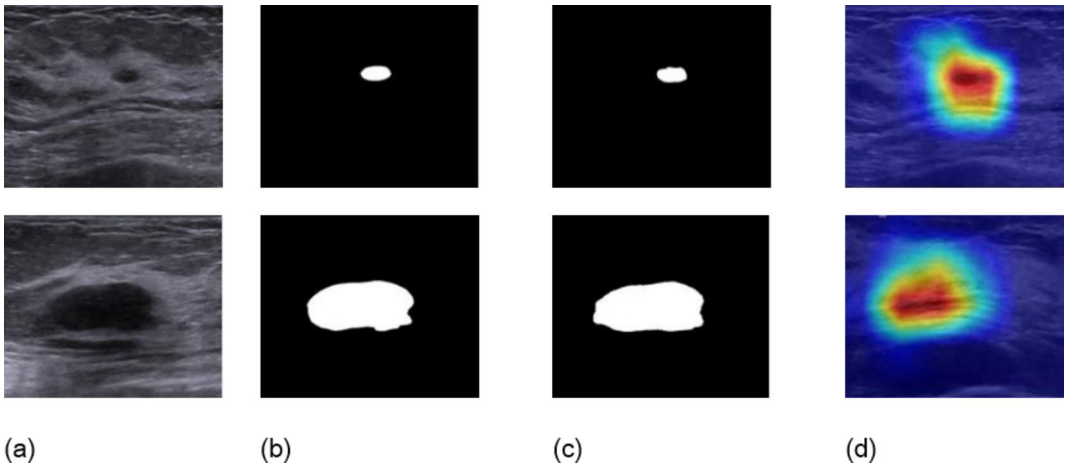

Fig. 9. Visualization results: (**a**) original image, (**b**) groundtruth mask, (**c**) predicted mask, (**d**) heat map.

Ref/year	Technique used	Dice score	Other performance parameters
2022[8]	Adaptive Attention U-Net	0.77	Precision = 0.79, Recall = 0.81, Specificity = 0.97
2022[9]	Residual cross-spatial attention-guided inception U-Net model	0.91	Precision = 0.94, Recall = 0.89
2023[10]	U2-MNet	0.82	Precision = 0.86
2023[11]	SU-Next	0.77	–
2022[12]	CNN	0.82	–
2021[16]	CNN	0.64	Jaccard Index = 0.56, Accuracy = 0.96
2020[23]	Selective Kernel U-Net	0.70	Accuracy = 0.95
2021[24]	Soft and hard attention mechanism	0.81	Accuracy = 0.95
2020[25]	Sub-region Pooling Network	0.84	Jaccard Index = 0.76
Proposed Model	**Attention-driven U-Net**	**0.85**	Precision = 0.97, Recall = 0.90, Jaccard Index = 0.73, Accuracy = 0.98, Specificity = 0.99

Table 5. State-of-art comparison on BUSI dataset.

in the field of breast cancer segmentation, highlighting developments that support the continuous improvement of state-of-the-art medical image processing methods.

Conclusions and future work

To sum up, the study presents a dependable and effective method for identifying breast cancer in ultrasound images using an Attention-driven U-Net model. The proposed approach performs extremely well in determining tumor boundaries using the Breast Ultrasound Images Dataset (BUSI), which consists of 780 images divided into malignant, benign, and normal instances. Among the key components of the proposed Attention U-Net architecture are the encoder, attention gate, and decoder blocks. They are designed specifically to replicate spatial elements, enhance information flow, and capture hierarchical aspects.

By offering a thorough approach and illustrating the Attention U-Net model's possible clinical relevance, this study makes a substantial contribution to the field of automated breast cancer segmentation techniques. The outstanding results and well-balanced assessment metrics highlight the importance of early detection for better patient outcomes and set the stage for upcoming developments in diagnostic interventions and medical imaging.

In future, the proposed Attention-driven U-Net model can be investigated on other imaging modalities such as Mammography, MRI or CT scans. Different imaging modalities have varying features discovering the model's performance to enhance its versatility. Also, integration of multi-modal information using a large-scale clinical validation diverse datasets from multiple medical institutions can be investigated for evaluating the model's reliability across different patient populations and imaging equipment.

Data availability

The data is publicly accessible as "Breast Ultrasound Images Dataset" and the link is: https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset.

Received: 25 March 2024; Accepted: 10 September 2024
Published online: 28 September 2024

References

1. Mnih, V., Heess, N., Graves, A. & Kavukcuoglu, K. Recurrent models of visual attention. *Adv. Neural Inf. Process. Syst.* **3**, 2204–2212 (2014).
2. Oktay, O. et al. Attention U-Net: Learning where to look for the pancreas. arXiv (2018).
3. Peng, C., Zhang, X., Yu, G., Luo, G. & Sun, J. Large kernel matters—Improve semantic segmentation by global convolutional network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1743–1751 (IEEE, 2017). https://doi.org/10.1109/CVPR.2017.189
4. Poudel, R. P. K., Lamata, P. & Montana, G. Recurrent fully convolutional neural networks for multi-slice MRI Cardiac Segmentation. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 83–94. https://doi.org/10.1007/978-3-319-52280-7_8 (2017).
5. Qin, Y. et al. Autofocus layer for semantic segmentation. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 11072 LNCS*, 603–611. https://doi.org/10.1007/978-3-030-00931-1_69 (2018).
6. Bhattacharya, T. et al. Applications of phyto-nanotechnology for the treatment of neurodegenerative disorders. *Materials* **15**(3), 804. (2022).
7. Sharma, B. & Koundal, D. Cattle health monitoring system using wireless sensor network: A survey from innovation perspective. *IET Wirel. Sens. Syst.* **8**(4), 143–151 (2018).
8. Chen, G., Li, L., Dai, Y., Zhang, J. & Yap, M. H. AAU-net: An adaptive attention U-net for breast lesions segmentation in ultrasound images. *IEEE Trans. Med. Imaging* (2022).
9. Punn, N. S. & Agarwal, S. RCA-IUnet: A residual cross-spatial attention-guided inception U-Net model for tumor segmentation in breast ultrasound imaging. *Mach. Vis. Appl.* **33**(2), 27 (2022).
10. Li, B., Su, B. & Zhang, X. U2-MNet: An improved neural network for breast tumor segmentation. In *Journal of Physics: Conference Series*, Vol. 2637, No. 1, 012003 (IOP Publishing, 2023).
11. Zhu, D. et al. *SU-Next: An image segmentation model of breast tumors based on U-next and attention mechanisms*. In *Journal of Physics: Conference Series*, Vol. 2575, No. 1, 012001 (IOP Publishing, 2023).
12. Podda, A. S. et al. Fully-automated deep learning pipeline for segmentation and classification of breast ultrasound images. *J. Computat. Sci.* **63**, 101816 (2022).
13. Huang, R. et al. Boundary-rendering network for breast lesion segmentation in Ultrasound images. *Med. Image Anal.* 02478 (2022).
14. Chen, G., Dai, Y. & Zhang, J. C-Net: Cascaded convolutional neural network with global guidance and refinement residuals for breast ultrasound images segmentation. *Comput. Methods Progr. Biomed.* 107086 (2022).
15. Tang, P., Yan, X., Nan, Y., Xiang, S. & Liang, Q. Feature pyramid non-local network with transform modal ensemble learning for breast tumor segmentation in ultrasound images. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* (2021).
16. Xue, C. et al. Global guidance network for breast lesion segmentation in ultrasound images. *Med. Image. Anal.* **70**, 101989 (2021).
17. Mishra, A. K., Roy, P., Bandyopadhyay, S. & Das, S. K. Breast ultrasound tumour classification:a machine learning—radiomics based approach. *Expert Syst.* **38**(7), e12713 (2021).
18. Xing, J. et al. Using bi-rads stratifications as auxiliary information for breast masses classification in ultrasound images. *IEEE J. Biomedical Health Inf.* **25**(6), 2058–2070. https://doi.org/10.1109/JBHI.2020.3034804 (2021).
19. Byra, M. Breast mass classification with transfer learning based on scaling of deep representations. *Biomed. Signal Process. Control* **69**, 102828. https://doi.org/10.1016/j.bspc.2021.102828 (2021).
20. Al-Dhabyani, W., Gomaa, M., Khaled, H. & Fahmy, A. Dataset of breast ultrasound images. *Data Brief.* **28**, 104863. https://doi.org/10.1016/j.dib.2019.104863 (2020).
21. Sulaiman, A. et al. Sustainable apple disease management using an intelligent fine-tuned transfer learning-based model. *Sustainability* **15**(17), 13228. (2023).
22. Sulaiman, A. et al. A convolutional neural network architecture for segmentation of lung diseases using chest X-ray images. *Diagnostics* **13**(9), 1651 (2023).
23. Byra, M. et al. Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network. *Biomed. Signal Process. Control* **61**, 102027 (2020).
24. Zhang, G. et al. SHA-MTL: Soft and hard attention multi-task learning for automated breast cancer ultrasound image segmentation and classification. *Int. J. Comput. Assist. Radiol. Surg.* **16**, 1–7 (2021).
25. Zhu, L. et al. A second-order subregion pooling network for breast lesion segmentation in ultrasound. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 160–170. (2020).

Author contributions

Adel Sulaiman: Conceptualization, Methodology, Software, Visualization, Investigation, Writing – review & editing. Vatsala Anand: Supervision, Data curation, Software, Visualization, Investigation, Writing – review & editing. Sheifali Gupta: Conceptualization, Methodology, Software, Visualization, Investigation, Writing – review & editing. Adel Rajab: Writing – review & editing, Methodology, Software. Hani Alshahrani: Software, Validation Writing – review & editing. Mana Saleh Al Reshan: Software, Validation, Writing – review & editing. Asadullah Shaikh: Software, Validation, Writing – review & editing, Supervision. Mohammed Hamdi: Software, Validation, Visualization, Supervision.

Funding

The authors are thankful to the Deanship of Graduate Studies and Scientific Research at Najran University for funding this work under the Easy Funding Program grant code NU/EFP/SERC/13/13.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.S.

Reprints and permissions information is available at www.nature.com/reprints.