Journal of Information Systems Engineering and Management
2025, 10(15s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

SURAKSHA: Segmenting Ultrasound Images using Real-Time Attention and Knowledge-based Structured Hybrid Architecture for Possible Breast Cancer Diagnosis

Keshav Kumar K.[1, 3*], N.V.S.L Narasimham[2] and A. Ramakrishna Prasad[3]
[1]*Department of Humanities and Mathematics, G. Narayanamma Institute of Technology and Science (for Women), Hyderabad-500 104, Telangana State, India.*
Orcid ID: https://orcid.org/0000-0002-9211-2960
[2]*Department of Humanities and Mathematics, G. Narayanamma Institute of Technology and Science (for Women), Hyderabad 500 104, Telangana State, India. Orcid ID: https://orcid.org/0000-0002-3572-9403*
[3]*Department of Mathematics, Jawaharlal Nehru Technological University, Hyderabad- 500 085, Telangana State, India.*
**Corresponding Author email- keshav.gnits@gmail.com*

ARTICLE INFO

Received: 30 Nov 2024

Revised: 20 Jan 2025

Accepted: 02 Feb 2025

ABSTRACT

Introduction: Breast cancer is one of the most common causes of death among women worldwide. Accurate and efficient diagnosis of breast cancer requires effective classification and segmentation of medical images. Though quite powerful, traditional deep learning mechanisms like Convolution Neural Networks (CNNs), Residual Neural Networks (ResNets), etc. are often computationally expensive, and resource-demanding. This calls for a Deep Learning approach to Constrained Quantization in Radiotherapy for Breast Cancer treatment by inheriting attention mechanisms.

Objectives: The primary objective of this research is to develop an efficient deep learning model for breast cancer diagnosis that enhances segmentation of medical images.

Methods: A novel deep learning architecture, SURAKSHA (Segmenting Ultrasound images using Real-time Attention and Knowledge-based Structured Hybrid Architecture) is proposed for possible Breast Cancer Diagnosis and identification of malignant tissues through the prediction of masks.

Results: The proposed model outperforms the state-of-the-art models by achieving accuracy (validation) of 93.01% on a lightweight cancer dataset, that is between 0.139% to 11.083% improvements from the existing models. Further, the possible variation in performance with changes in the dimensions of the select patch is studied.

Conclusions: The study presents an efficient deep learning-based segmentation to accurately detect malignant breast tissues, aiming to improve the effectiveness of radiotherapy.

Keywords: Breast Cancer, Computer Vision, Ultrasonography, Attention Mechanism

1. INTRODUCTION

Breast cancer is a major public health challenge, and is one of the most prevalent forms of cancer and a leading cause of mortality among women worldwide (refer to Table 1). Patient Outcomes depends majorly on timely detection and diagnosis of malignancy in the mammary tissues, necessitating the need for effective classification and segmentation models for medical images. Among different imaging methods, breast ultrasonographic images (refer to Fig. 1) are often found as a valuable tool due to their accessibility, cost-effectiveness, and capability to provide real-time imaging. Despite these positives, the interpretation of ultrasound images is very broadly subjective. It relies entirely on the expertise of medical professionals supervising the patient, which leads to variability in diagnostic accuracy, adding redundancy to the diagnosis process. Following is the problem statement; we are willing to address through this research.

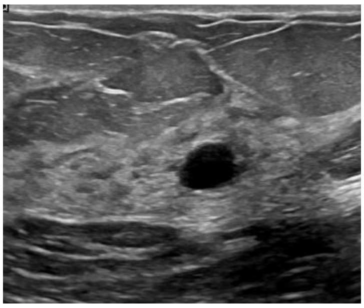
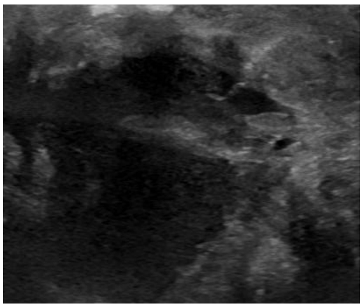
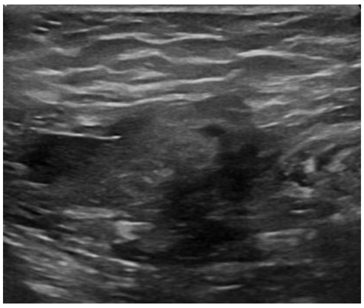
Problem Statement: Constrained Quantization in Radiotherapy of Breast Cancer treatment.

Proposed Approach: In Radiotherapy, identification (rather than quantification) of the target site is a mandatory preliminary. While, optimization techniques like Intensity-Modulated Radiation Therapy (IMRT), Q-learning, Volumetric Modulated Arc Therapy (VMAT), etc. are successful to some extent, it calls for the need of a more accurate, yet computationally feasible solution. We propose to make use of Deep Learning based Segmentation to identify malignant breast tissue, and thereafter perform radiotherapy.

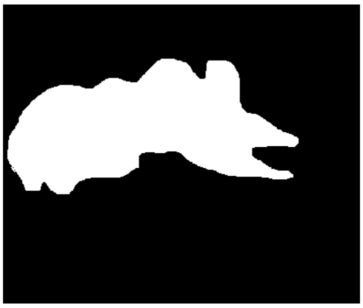
Table 1: Here are Approximate Mortality Figures for the World, divided into five major regions: Latin America, Northern America, Asia, Oceania, and Europe. The Figures are obtained from the Global Cancer Observatory (https://gco.iarc.who.int/en).

Region	**Mortality (as of 2023)**	**Mortality (as of 2022)**	**Mortality (as of 2021)**
Latin America	≈ 68,000	≈ 48,000	≈ 46,000
Northern America	≈ 43,170	≈ 43,700	≈ 43,600
Asia	≈ 440,000	≈ 430,000	≈ 411,000
Oceania	≈ 693,000	≈ 670,000	≈ 619,200
Europe	≈ 138,000	≈ 140,000	≈ 138,000

Recent technological advances in deep learning and computer vision have greatly supported the medical imaging field, providing robust models and architectures for automated image analysis. Traditional architectures, such as Convolutional Neural Networks (CNNs) and Residual Neural Networks (ResNets), have shown promising results in various image classification tasks, but are computationally inefficient [1]. Moreover, the lack of interpretability of the results in traditional deep learning-based models lacks clinical acceptance, as healthcare practitioners seek tools that not only provide accurate (mostly), and understandable predictions.

Benign Cancer Malignant Cancer Normal

N/A

Mask (Benign) Mask (Malignant) Mask Not Applicable

Figure 1: Variants and Carcinogenicity of Breast Ultrasound Images from Subjects ages 25 to 75. For each instance, the corresponding mask, specifically marking the affected region, has been shown.

Existing literature includes several works incorporating deep learning models for medical image analysis. Pattnaik *et al.* in their contribution, "*Breast cancer detection and classification using metaheuristic optimized ensemble* *extreme learning machine*", proposed a meta-heuristic algorithm for the detection (segmentation, and classification) of breast cancer from mammogram images [2]. It outperforms state-of-the-art ML models and expert supervision-based analysis. Behera *et al.* suggested the use of UNet and Residual UNet (ResUNet) architectures for the segmentation of skin lesions (malignant melanoma), and achieved better accuracy, Dice coefficient, Jaccard index, sensitivity, and specificity than the s-o-t-a in their work, "*Melanoma skin cancer detection using deep learning-based lesion segmentation*" [3]. Kaladevi *et al.* in their research, "*Morpho-contour exponential estimation algorithm for predicting breast tumour growth from MRI imagery*", introduced Morpho-Contour Exponential Estimation (MoCEE) for the identification of cancerous tissues (segmentation task) [4], and achieved 92.22% accuracy (**state-of-the-art**) on a lightweight ultrasound image dataset. It is evident from the findings of these researches that, deep learning models can achieve much higher accuracy than which is observed by expert supervision, but are computationally expensive, and have poor interpretability.

To address these limitations, *Attention Mechanisms* have gained attention as a means to enhance the performance of deep learning models. For Example, Works by Bandaru *et al.* proposed a fusion of Swin Transformer and Attention Mechanisms for the classification of liver cancer and achieved 99.29% accuracy [5]. By enabling the model to focus on relevant features in the input data, attention mechanisms improve accuracy while reducing computational overhead. Mohanty *et al.* introduced Attention Mechanisms with a cluster of 4 Convolutional Neural Network for MRI-based brain tumour classification and achieved better accuracies than several s-o-t-a [6].

This research introduces **SURAKSHA** (**S**egmenting **U**ltrasound images using **R**eal-time **A**ttention and **K**nowledge-based **S**tructured **H**ybrid **A**rchitecture), an innovative framework designed to leverage real-time attention and structured hybrid architectures for the segmentation of breast ultrasound images. The proposed model outperforms the state-of-the-art models by achieving accuracy (validation) of 93.01% on the lightweight cancer dataset. **SURAKSHA** gives the flexibility to choose the dimensionality of image patches. Therefore, the possible variation in performance with changes in the dimensions of the select patch is also studied, and therefore reported. On a gist, the main contributions of the work can be enumerated as follows,

1. Proposition of **SURAKSHA** (**S**egmenting **U**ltrasound images using **R**eal-time **A**ttention and **K**nowledge-based **S**tructured **H**ybrid **A**rchitecture), which outperforms the state-of-the-art model (92.22% validation accuracy) by achieving a segmentation accuracy (validation) of 93.01% on the lightweight cancer dataset.
2. Combined both the **Traditional UNet Architecture** with the state-of-the-art **Attention Mechanisms** for better segmentation.
3. Added flexibility options for **SURAKSHA** to choose the dimensionality of image patches under consideration (64 × 64, 128 × 128, 256 × 256) as per the requirements.

2. Proposed Model- SURAKSHA

This section proposes **SURAKSHA** (**S**egmenting **U**ltrasound images using **R**eal-time **A**ttention and **K**nowledge-based **S**tructured **H**ybrid **A**rchitecture), and discusses the hybrid architecture of the models in detail. On a high level, it can be thought of as a hybridized connection of the traditional **UNet Model** [7] with stacked **Attention Mechanisms** [8]. Figure 2 gives a pictorial depiction of the **SURAKSHA** pipeline.

To begin, the model feeds on a training set comprising a set of ultrasound image data points of dimensions($H \times W \times n_c$), with the ultrasound images being of ($H \times W$) pixels, and n_c color channels (e.g. 1 for Grayscale, 3 for RGB, 4 for CMYK, etc.). The input ultrasound images are then fed to a series of 4 Encoder Blocks (EBs). The input images possess a cluster of features collectively, which cannot be distinguished by the normal human eyes. Thus, comes to play the Encoder Blocks, which have been designed specifically to extract hidden features from the images in a hierarchical pattern. The reason behind 4 progressive Encoder Blocks is to enable the model to capture higher-level, and complex features, which are specifically the regions of interest for the research. Further, in each Encoder Block, there are MaxPool2D Layers that aims to **down-sample the images successively**. This reduces the spatial dimensions, allowing the subsequent neuronal layer to focus on the prominent features besides minimizing computational complexity. Each Encoder Block **captures larger receptive fields** (i.e., it integrates information from a larger area of the input image) with respect to the previous Encoder Block. Figure 3 (a) shows the detailed implementation of the molecular Encoder Block.

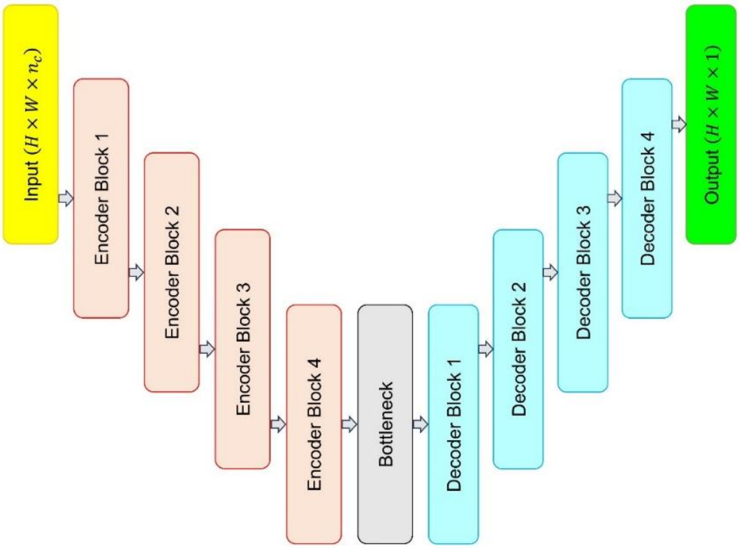
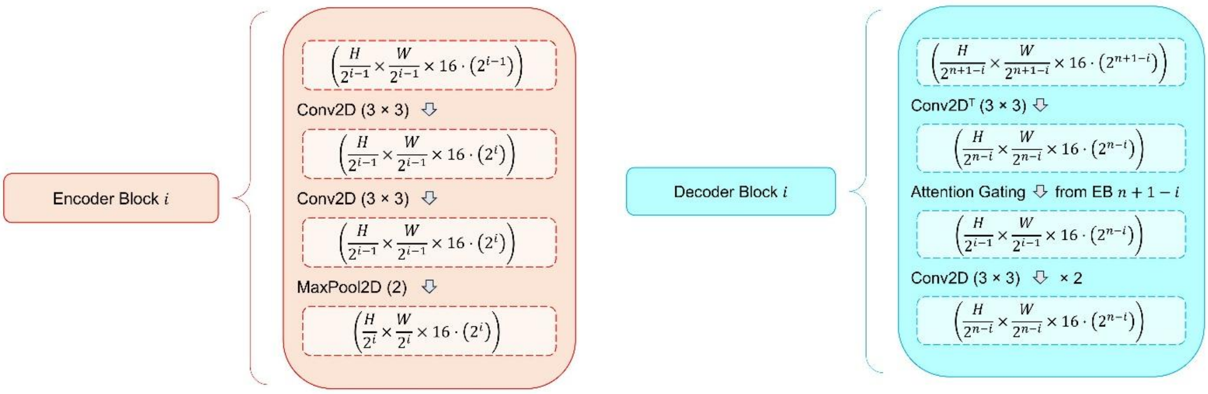


Figure 2: Pipeline of the **SURAKSHA** (**S**egmenting **U**ltrasound images using **R**eal-time **A**ttention and **K**nowledge-based **S**tructured **H**ybrid **A**rchitecture) Architecture. The detailed implementation of the Encoder Blocks (EBs), and the Decoder Blocks (DBs) have been presented in Figure 3.

Following the EBs, we have a Bottleneck layer that operates at the lowest image resolution thus capturing the most complex, abstract, and high-level features. The respective EBs are able to achieve a local contextual understanding of the image, but the Bottleneck layer is able to achieve both the **Global and local contextual understanding**. Further, it is a pathway to pass the learned information (bits) to the Decoder Blocks (DBs).

The next phase in the **SURAKSHA** pipeline, is the Decoding phase consisting of a sequence of four Decoder Blocks (DBs), that reconstruct the image from the abstract features learned by the model in the Encoding phase. Each DB **up-samples the images successively** from the learned feature maps in the Encoding Phase, and the Bottleneck to subsequently reach the dimensions of the input image of $(H \times W)$ pixels. The Decoder Blocks further **combine with the attention mechanism** through **skip-connections** [9], allowing the model to concentrate on the feature relevance learned by the congruent Encoder Block. After the final upsampling stage, the decoder passes the feature maps through a Conv2D layer (two-dimensional Convolution Layer) to predict the final output, i.e. mask for the region of interest of dimension$(H \times W \times 1)$. Figure 3 (b) shows the detailed implementation of molecular Decoder Block. In the Training Phase, the predicted mask is then compared with the actual mask, and a binary cross entropy loss is considered to update the weights and biases influencing the mask prediction.



(a) Implementation of i^{th} Encoder Block (EB) **(b)** Implementation of i^{th} Decoder Block (DB)

Figure 3: Detailed Implementations of the i^{th} Encoder Block (EB), and i^{th} Decoder Block (DB)

3. EXPERIMENTAL SETUP

This section discusses about the Experimentation Process used to evaluate the proposed **SURAKSHA** Model. The Validation Phase of the model involves a few metrics related to Segmentation as discussed in Subsection 3.2. The observed results from benchmarking (evaluation) of the model on the mentioned Experimental Setup are presented in Subsection 3.3.

3.1. Dataset

Al-Dhabyani *et al.* in their work, "*Dataset of breast ultrasound images*", provided a light-weight (780 image samples) image set of average dimension 500 × 500 pixels from 600 patients aged between 25 and 75 years old, collected from the *Baheya Hospital for Early Detection & Treatment of Women's Cancer, Cairo, Egypt* [10]. For the research conducted in this paper, we considered the dataset from Al-Dhabyani *et al.* Two main reasons assisted such a choice for us,

1. The dataset was well documented and open-sourced, with a well-defined naming convention making it easier to implement the **SURAKSHA** Model, for example, for each input (say us_image.png), the respective mask was named by adding the token "_mask" to the end of the filename, i.e., us_image_mask.png
2. Generally, most Ultrasound (US) dataset considers 2 states of malignancy – "*negative*", and "*positive*". This dataset considers 3 different states of Breast Cancer – "*malignant*", "*benign*", and "*normal*" with their respective feature masks.

Figure 1 gives an instance of the dataset comprising the image along with its ground truth mask.

3.2. Metrics

Since the work deals with Segmentation Problem, and Image Segmentation, the most Common Metrics are Accuracy (observed during Validation), Intersection Over Union (IoU) Score and Dice Coefficient. Thus, to evaluate the performance of the proposed **SURAKSHA** Model with respect to the prevalent Deep Learning based Medical Image Segmentation models, the following metrics have been used.

1. **Training Accuracy**, which is the area (or percentage, if needed) of the correctly identified segmentation region with respect to the total region on the training dataset. Mathematically, it is defined as,

$$\text{Accuracy}_{\text{train}}(\%) = \frac{\#\text{ Correctly Classifed Pixels}}{\#\text{ Total Pixels}} \times 100$$

2. **Testing Accuracy**, which is the area (or percentage, if needed) of the correctly identified segmentation region with respect to the total region on the testing dataset. Mathematically, it is same as the Training Accuracy, the only difference being the collection of datasets it is worked upon.
3. **Intersection Over Union (IoU)**, which in layman's terms defines the overlapping area between the predicted segment (s_p) and the ground truth (s_g). Mathematically, it is defined as,

$$\text{IoU} = \frac{|s_p \cap s_g|}{|s_p \cup s_g|}$$

4. **Dice Coefficient**, which is another measure to define the overlapping area between the predicted segment (s_p) and the ground truth (s_g). Mathematically, it is defined as,

$$\text{DC} = \frac{2 \cdot |s_p \cap s_g|}{|s_p| + |s_g|}$$

5. **Precision**, evaluates how many of the predicted pixels (that the model identified as belonging to the target class, i.e., True Positive) are actually correct. Mathematically, it is defined as,

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

6. **Recall**, evaluate how many of the actual pixels (that belong to the target class) are correctly identified by the model as positive instances. Mathematically, it is defined as,

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

Additionally, the time taken by the models to train themselves (in seconds) on the given dataset has also been reported as a superficial metric to further compare the models having similar responses for the metrics defined earlier.

3.3. Experimental Results

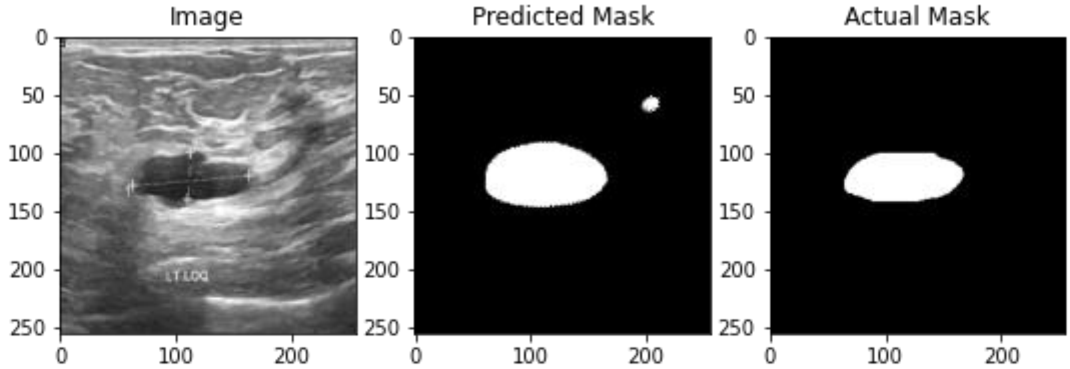
This subsection is directed towards presenting the achieved results using the Dataset in Subsection 3.1 as a benchmark. Besides comparing the proposed **SURAKSHA** Model with respect to the prevalent Deep Learning based Medical Image Segmentation models (namely, UNet [7], UNet++ [11], SegNet [12], RefineNet [13], and FCN [14]), 3 different variants of the **SURAKSHA** Model are considered – C1, C2, and C3 respectively, corresponding to different input sizes. C1 being the least computationally expensive model considers the **input images of size 64 by 64**. Thus, passing through the respective Encoder Blocks, the images are condensed to 32 by 32, 16 by 16, 8 by 8, and finally 4 by 4 respectively. C2 being the mild computationally expensive model considers the **input images of size 128 by 128**. Thus, passing through the respective Encoder Blocks, the images are condensed to 64 by 64, 32 by 32, 16 by 16, and 8 by 8 respectively. C3 being the most computationally expensive model considers the **input images of size 256 by 256**. Thus, passing through the respective Encoder Blocks, the images are condensed to 128 by 128, 64 by 64, 32 by 32, and 16 by 16 respectively.

The respective comparative analysis between all the models is presented in Table 2. From the presented results, it can be observed that the proposed **SURAKSHA** Model offers a **0.139% to 11.083% improvement** in validation efficiency.

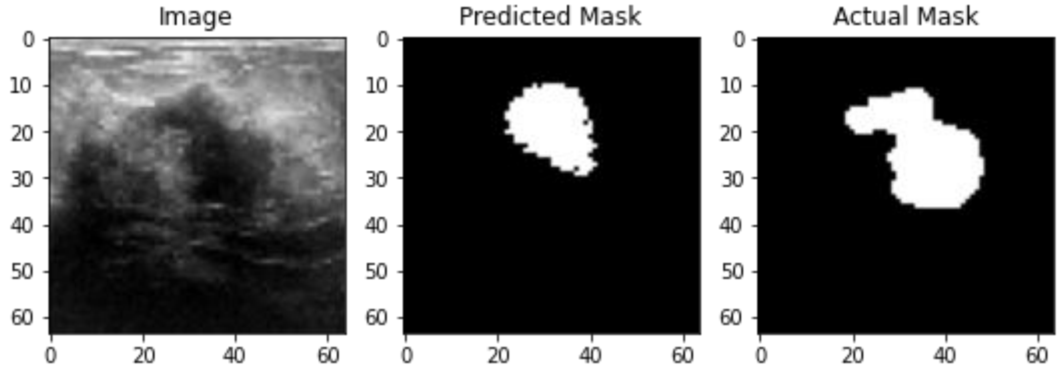
Metrics	**UNet** [7]	**UNet++** [11]	**SegNet** [12]	**RefineNet** [13]	**FCN** [14]	**SURAKSHA** (proposed)		
						C1	**C2**	**C3**
Accuracy (Training)	0.9573	0.9541	0.9553	0.9566	**0.9588**	0.9481	0.9534	0.9412
Accuracy (Validation)	0.8373	0.9037	0.9288	0.9048	0.9170	0.8992	0.9183	**0.9301**
Intersection Over Union	0.3090	0.3833	0.3508	0.3528	0.1604	0.3841	0.3841	**0.4487**
Dice Coefficient	0.4721	0.5542	0.5194	0.5215	0.2764	0.5551	0.3376	**0.6195**
Precision	0.3363	0.4850	0.6846	0.4795	**0.9325**	0.4552	0.7991	0.6283
Recall	**0.7916**	0.6464	0.4184	0.5716	0.1622	0.7111	0.2140	0.6110
Time to Train (seconds)	701.78	433.99	247.87	201.02	**85.63**	338.44	640.70	674.51

Table 2: Experimental Results of the Deep-Learning-Based Segmentation Models, when experimented on Breast Ultrasound Images. The results show that the proposed **SURAKSHA** Model outperforms other models in terms of its Accuracy (Validation), IoU Score, and Dice Coefficient, which are the most prominent metrics for evaluating any Segmentation Model. However, the best performers under each metrics have been **highlighted**.

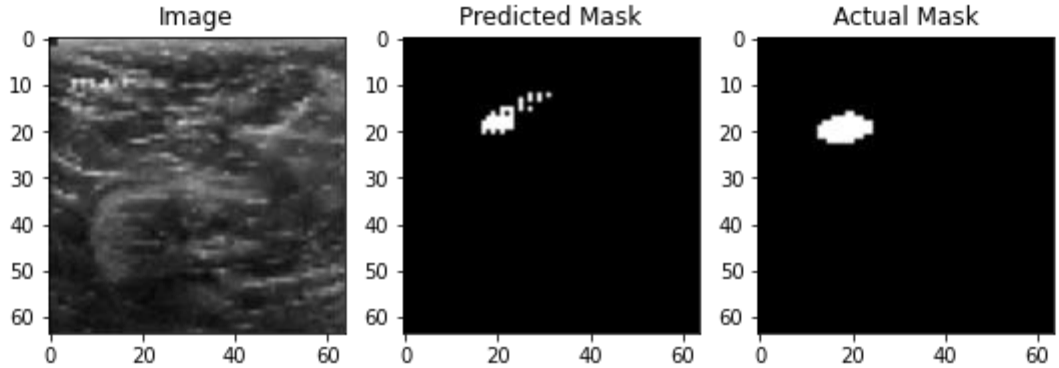
Figure 4 gives the US Image, corresponding predicted and actual mask for the top 3 performing models based on their validation accuracy – **SURAKSHA (C3)**, **SegNet**, and **SURAKSHA (C2)** respectively.



(a) Visualization of the US Image, along with the masks for the **SURAKSHA (C3)** Model. **(Best Performer)**



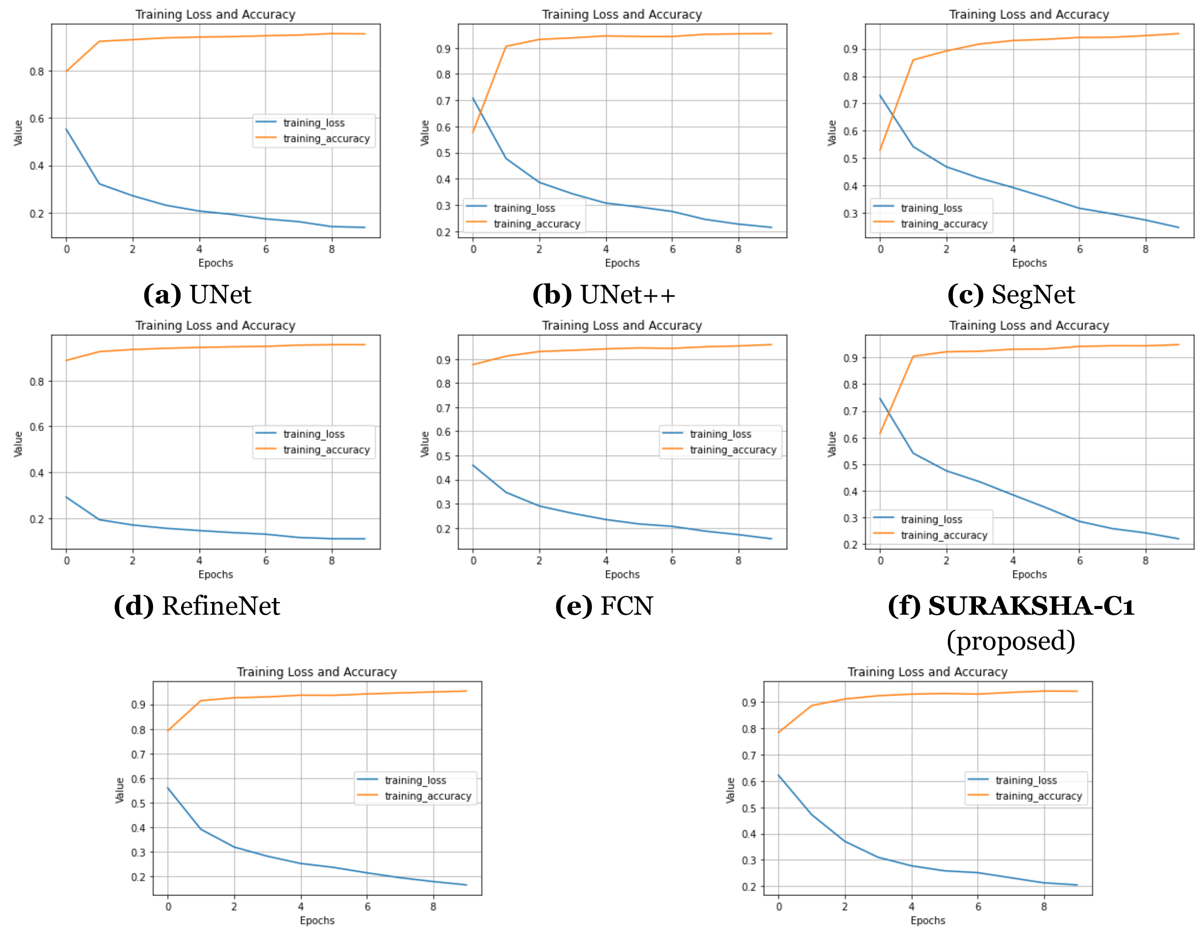
(b) Visualization of the US Image, along with the masks for the **SegNet** Model.



(c) Visualization of the US Image, along with the masks for the **SURAKSHA (C2)** Model.

Figure 4: Comparative analysis between the top 3 performers (state-of-the-art) models for segmentation task especially for the light Breast Ultrasound dataset from the works of Al-Dhabyani *et al.*

Figure 5 shows the epoch-wise variation in accuracy (training), and the loss for all the 8 models discussed, and mentioned in Table 2.



(a) UNet **(b)** UNet++ **(c)** SegNet

(d) RefineNet **(e)** FCN **(f) SURAKSHA-C1** (proposed)

(g) SURAKSHA-C2 (proposed) **(h) SURAKSHA-C3** (proposed, **best**)

Figure 5: Epoch wise variation in Training Accuracy and Loss (Binary Cross entropy) for each of the eight different deep learning models experimented on in the research. Models like RefineNet and FCN starts with very high training accuracy, but lacks in the validation phase. This symbolizes possible overfitting with the mentioned models. It is to be noted that all of our proposed models start from comparatively lower training accuracy, and then learn the patterns to get a higher accuracy in subsequent epochs, **which is a desirable signature for good learners.**

4. DISCUSSION AND LIMITATIONS

SURAKSHA (**S**egmenting **U**ltrasound images using **R**eal-time **A**ttention and **K**nowledge-based **S**tructured **H**ybrid **A**rchitecture) is a hybridization of the traditional UNet Architecture with state-of-the-art attention mechanisms. This improvised version of the traditional UNet Architecture not only helps to achieve better accuracy (in terms of validation), but is also less computationally expensive. For the implementation, we considered Keras as a Backend, and executed the models on a free access cloud-based notebook environment provided by the Google Colaboratory. The codes and supplementary materials will be made available upon reasonable request. We are planning to open-source the materials post-publication. Hereby, since we are nearing the conclusion part of this research, we would like to answer two most obvious questions, **a) Why SURAKSHA?** and **b) When SURAKSHA?** Answering each of these in chronological order,

a) **SURAKSHA** is a possible alternative to the traditional segmentation algorithms, not only because it achieves higher accuracy than several other deep-learning based models, but also because it achieves it in much lesser time compared to the preliminaries. Quantitatively, **SURAKSHA** achieves a 0.139% to 11.083% improvement in results.

b) **SURAKSHA** would be a good alternative to the traditional segmentation algorithms, if the dataset on which the model is trained is a lightweight one (with comparatively lesser number of sample points).

Though, technically and scientifically appealing, we would like to address a few shortcomings of **SURAKSHA**.

1. **SURAKSHA** have only been tested on breast ultrasound images, and we are unsure of its' performance on other medical segmentation tasks.
2. Though, **SURAKSHA** is able to achieve good accuracy, and is computationally efficient than many MLP based Models like UNet, UNet++, etc., but there exits, several models like SegNet, RefineNet etc, which got their results in much lesser time.
3. There are some state-of-the-art technologies in Object Detection tasks that use a dynamic patch detection mechanism during run-time**.** **SURAKSHA** doesn't possess the dynamical patch selection mechanism, rather the patch size needs to be hard-coded in the compile time.

5. CONCLUSION

This research introduces **SURAKSHA**, a hybrid deep learning model specifically developed to enhance segmentation accuracy in breast ultrasound images. The research addresses a key issue in breast cancer radiotherapy: constrained quantization. Identifying and accurately quantifying the cancerous tissue is critical for effective treatment. While existing techniques like IMRT, Q-learning, and VMAT have had some success, they often fall short of achieving both the necessary accuracy and computational efficiency in practice.

The proposed approach utilizes deep learning-based segmentation to accurately detect malignant breast tissues, aiming to improve the effectiveness of radiotherapy. With a **validation accuracy of 93.01%**, **SURAKSHA** significantly outperforms other methods, showing **improvements ranging from 0.139% to 11.083%**. One of the standout features of this model is its flexibility, as it allows for different patch sizes without compromising accuracy, depending on computational resources. Additionally, the incorporation of attention mechanisms improves the model's ability to focus on critical features for better segmentation. In essence, **SURAKSHA** offers a practical and more reliable solution for breast cancer treatment, combining accuracy and computational feasibility. Moving forward, we will work on optimizing this model for real-time clinical use, which could have a meaningful impact on improving outcomes for patients undergoing radiotherapy.

Conflict of interest:

The authors declare no conflict of interest.

Funding: No funds, grants, or other support was received.

Financial Interests: The authors declare they have no financial interests.

REFERENCES

[1] Dutta, A., Lakshmanan, K., Sivakumar, P., Ramamoorthy, A., Kumar, P. K., & Harshith, J. (2023, August). Forecasting of Cyclone Severity using Intelligent NeuralNets subjecting to Infrared Imaging. In *2023 9th International Conference on Smart Computing and Communications (ICSCC)* (pp. 268-273). IEEE.https://doi.org/ 10.1109/ICSCC59169.2023.10335019.

[2] Pattnaik, R. K., Siddique, M., Mishra, S., Gelmecha, D. J., Singh, R. S., & Satapathy, S. (2023). Breast cancer detection and classification using metaheuristic optimized ensemble extreme learning machine. *International Journal of Information Technology*, *15*(8), 4551-4563.https://doi.org/ https://doi.org/10.1007/s41870-023-01533-y

[3] Behera, N., Singh, A. P., Rout, J. K., & Balabantaray, B. K. (2024). Melanoma skin cancer detection using deep learning-based lesion segmentation. *International Journal of Information Technology*, 1-16.https://doi.org/10.1007/s41870-024-02004-8

[4] Kaladevi, P., Kanimozhi, N., Nirmala, B., & Sivasankari, R. (2024). Morpho-contour exponential estimation algorithm for predicting breast tumor growth from MRI imagery. *International Journal of Information Technology*, 1-16. https://doi.org/10.1007/s41870-024-01803-3

[5] Bandaru, S. C., Mohan, G. B., Kumar, R. P., & Altalbe, A. (2024). SwinGALE: fusion of swin transformer and attention mechanism for GAN-augmented liver tumor classification with enhanced deep learning. *International Journal of Information Technology*, 1-19.https://doi.org/10.1007/s41870-024-02168-3

[6] Mohanty, B. C., Subudhi, P. K., Dash, R., & Mohanty, B. (2024). Feature-enhanced deep learning technique with soft attention for MRI-based brain tumor classification. *International Journal of Information Technology*, *16*(3), 1617-1626.

[7] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18* (pp. 234-241). Springer International Publishing.

[8] Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.

[9] Wu, D., Wang, Y., Xia, S. T., Bailey, J., & Ma, X. (2020). Skip connections matter: On the transferability of adversarial examples generated with resnets. *arXiv preprint arXiv:2002.05990*.

[10] Al-Dhabyani, W., Gomaa, M., Khaled, H., & Fahmy, A. (2020). Dataset of breast ultrasound images. *Data in brief*, *28*, 104863.

[11] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4* (pp. 3-11). Springer International Publishing.

[12] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, *39*(12), 2481-2495.

[13] Lin, G., Milan, A., Shen, C., & Reid, I. (2017). Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1925-1934).

[14] Sun, W., & Wang, R. (2018). Fully convolutional networks for semantic segmentation of very high resolution remotely sensed images combined with DSM. *IEEE Geoscience and Remote Sensing Letters*, *15*(3), 474-478.