ISSN: 0976-9102 (ONLINE)
DOI: 10.21917/ijivp.2025.0504

A COMPARATIVE STUDY OF DEEP LEARNING-BASED SEGMENTATION TECHNIQUES: U-NET, SEGNET, AND DEEPLABV3 FOR BREAST CANCER DETECTION BY THERMAL IMAGING

M. Kanaga[1], K. Saipriya[2] and N. Madurai Meenachi[3]
Department of Atomic Energy, Indira Gandhi Centre for Atomic Research, India

Abstract

Image segmentation is vital for a wide range of computer vision applications, where precise delineation of regions within an image is essential. The breast thermal images acquired are segmented manually using various tools such Photoshop, label box. Images can also be segmented semi automatically using point-based annotation, a bounding box, etc. As the, semi-automatic methods are time-consuming with large datasets, a fully automatic segmentation model is required to provide a better solution. To address this, pretrained models with various backbone networks can be used with a few manually annotated masks to segment the images. The three different architecture models, namely U-net, SegNet, and a pretrained DeepLabv3 were analyzed, using the datasets holding 1000 images and their corresponding mask (ground truth). The three models have been compared based on accuracy, pixel accuracy, dice score and Intersection over Union. U-Net yielded better results, achieving a dice coefficient of 96 %, and it required less training time than the SegNet, which managed a dice score of 93%. DeepLabv3 showed relatively lower performance with a Dice Coefficient of 89%.

Keywords:

U-Net, SegNet, Threshold, DeepLabV3, Segmentation

1. INTRODUCTION

Breast cancer has the highest mortality rate of 12.41 per lakh population of all cancer types in Indian women. Hence it is highly required to create awareness on the early detection through self-examination and through regular check up on women attaining the age of 50. The early diagnosis of breast cancer significantly increases the chances of survival and the likelihood of a successful cure [1]. Hence, effective screening techniques are being continuously evolved to detect breast cancer at early stage itself. In India, the most common screening technique is mammogram. As the number of radiologists is relatively less in India than in developed countries [2] for cancer screening, the thermal breast screening combined with artificial intelligent will be an alternative technique. Various types of Machine Learning and Deep Learning model can be used for classification of images for proper classification of breast thermal images, the Region of interest (ROI) has to be identified and subsequently segmented for achieving better accuracy [3].

Segmentation of ROI is a critical step in image processing which is the prime step in determining the model's accuracy. As regards to breast thermal images, the ROI in images are small with more vital information. In these images, the usage of segmentation techniques will locate the ROI and omit unwanted information, hence, focusing on the ROI will decrease the computation time and increase the accuracy [4]. In thermal breast images, as the temperature variation between benign and malignant is minimum, identification of ROI is challenging [5].

2. LITERATURE SURVEY

The segmentation techniques can broadly be categorized into three categories: manual, semi-automatic, and fully automatic. The automatic segmentation techniques have been elaborated in detail in this literature survey.

In an automatic segmentation method, the ROI is identified using predefined parameters and protocols. The protocol involves maintaining fixed distance between patient and camera for imaging. The acquired image is then cropped based on predefined coordinates along the height, thereby identifying the required ROI without utilizing any segmentation algorithm [6].

A fully automated segmentation algorithm relies on thresholding and combining multiple traditional image processing techniques, which may also involve minimal manual intervention. Images have been processed through Canny Edge Detection, followed by the Hough Transform to identify connected edges for the breast boundary estimation. A circular curve is then used for region isolation [7], [8].

In order to completely eliminate the need for manual intervention, Deep Learning model has been used for fully automated segmentation. Some of the Deep Learning models are U-Net, MultiResU-Net (extension of U-Net) [9], [10], FCN, SegNet , DeepLabv3.

Ange Lou et al. [9] compared the convolutional and deconvolutional neural networks, U-Net, and MultiResUnet architectures for segmenting a thermal breast dataset consisting of 450 images. The acquired images were preprocessed, smoothed, and manually cropped before being input into the algorithms. U-Net achieved an accuracy of 89.8%. In this work the author used Tanimoto similarity metric instead of IoU for evaluating the algorithm

In another study by Dalmia et al. [11], various deep learning networks for tumor detection is compared, which includes VGGNet, InputCascadeCN, U-Net, and V-Net. In this study, U-Net achieved a high accuracy of 99.5%, but its Dice coefficient (76%) and Jaccard Index (61.3%) were relatively lower compared to other models. In contrast, V-Net outperformed U-Net in segmentation quality, with a Dice coefficient of 79% and IoU of 66.2 %.

Vianna et al. [12] evaluated the performance of the ultrasound breast images segmented using SegNet and U-Net and achieved Dice coefficients of 86.3% and 81.1%, respectively. However, Alqazzaz et al. [13] highlighted that the segmentation of images using SegNet is time-consuming during the training phase and algorithms have also been used for directly segmenting the tumor region in addition to ROI. In another study, Ahmed et al. [14] investigated the segmentation of breast tumor imaged by mammography using DeepLabv3 and achieved a Mean Average Precision (mAP) of 72% .

On reviewing the literature and analyzing the pros and cons of various segmentation methods U-Net, SegNet, and DeepLabv3 have been choosen for comparison based on their merits. This article highlights the study and comparison of segmentation models using U-Net, SegNet, and DeepLabv3 for thermal breast images.

3. METHODOLOGY ADOPTED

3.1 HARDWARE AND SOFTWARE SPECIFICATION

The Python programming language was used to implement the model, with Keras and TensorFlow libraries from the Anaconda package. The development of the model was done within the Jupyter Notebook Integrated Development Environment, while the front-end interface was designed using Visual Studio Code, with Flask employed to facilitate web application deployment.

3.1.1 Computational Setup:

In order to study the computational time to train the model, all three models are trained in two different systems named as Machine 1 and Machine 2, with the following specifications: Machine 1 setup is given below:

- **Processor:** Intel(R) Core(TM) i7-10700 CPU @ 2.90GHz
- **RAM:** 16.0 GB
- **Processor Architecture:** x64-based processor
- **Operating System:** Windows 11 Pro

Machine 2 setup is given below

- **Processor:** Intel(R) Xeon(R) CPU E5-2687W v0 @ 3.10 GHz (Dual processors)
- **RAM:** 96.0 GB
- **Operating System:** Windows 10 Pro (64-bit)
- **Processor Architecture:** x64-based processor

It is observed that, during the training phase, Machine 1 encountered extended computation time with occasional system crash. The extended time is attributed to the increased computational demand during the training process. In contrast, Machine 2 completed the tasks without such issues. The computational time for both machines is compared for each model to assess performance differences as shown in Table.2. The total time taken for training the model is calculated using the formula as in Eq.(1):

$$\text{Total time taken for training} = \text{Number of epochs} * \text{Time taken for one epochs} \quad (1)$$

3.1.2 Data Acquisition:

The dataset for understanding the segmentation model is taken from Figshare. The data set contains 1000 thermal breast images with a 224*224 pixels resolution, and the mask of 1000 images. This data set which is used for evaluating the metrics is split into three sets: training (70%), validation (10%), and testing (20%).

3.2 U-NET

U-Net is encoder and decoder network architecture for fast and precise segmentation of images [12]. Its contracting path (encoder) captures context and reduces the spatial dimensions. The expansive path (decoder) increases the spatial dimensions to restore the original image size and provides pixel-wise classification. This model uses batch normalization after each convolutional layer and dropout layers. This helps to prevent over fitting and stabilize the training process [3].

The encoder path of U-Net [10, 11] architecture consists of four convolutional blocks. Every convoluted block has two Conv2D layers followed by a max pooling layer. The number of filters in these blocks increases as the network deepens, with 64 filters in the first block, 128 in the second, 256 in the third, and 512 in the fourth block. This increasing number of filters allows the network to capture more complex features at each stage.

The bottleneck layer, situated between the encoder and the decoder, represents the deepest layer of the network. Bottleneck layer consists of two Conv2D layers with 1024 filters each, where the highest level of feature extraction occurs. The bottleneck layer captures the most abstract and global features of the input image, serving as the foundation for the upsampling in the decoder.

The decoder has three convolutional blocks. Each block includes a Conv2D Transpose layer, which performs upsampling, followed by two Conv2D layers to further refine the features. These blocks progressively increase the resolution of the feature maps and use skip connections to combine high-level features from the encoder path, ensuring accurate reconstruction of the original image.

Adam optimizer with a learning rate of $1e^{-4}$, a sigmoid activation function in the output layer, and Binary Cross-Entropy (BCE) as the loss function have been used for training the model.

3.3 DEEPLABV3

DeepLabv3 is a pretrained model developed by Google with the predecessor of DeepLabv1 and DeepLabv2. DeepLabv3 is trained with different backbone networks like imagnet, mobilenet and densenet. This model accepts RGB images with size of 224*224. The U-Net and SegNet accepts gray scale images but DeepLabv3 accepts RGB. The dataset grey scale image is converted into RGB. The DeepLabv3 is extremely accurate when it comes to multi-scale segmentation [15].

While in masks preprocessing, the ground truth (mask) is loaded and resized to match the dimensions of the images. The masks are converted to binary values, where the foreground is labeled as 1 and the background as 0. This binary labeling is essential for segmentation tasks, as it enables to identify ROI of the image [16].

DeepLabv3 has been used along with various pretrained models. In this article DeepLabv3 framework uses DenseNet201 model which has a backbone network on ImageNet that serves as the feature extractor for this architecture. The output from the DenseNet201 is then passed via a Conv2D layer with softmax activation to produce class predictions for each pixel. As it is binary segmentation, the last layer has softmax activation function.

To restore the original image size, an UpSampling2D layer is inserted. Sparse categorical cross-entropy is appropriate when labels are represented by integer values. Sparse categorical cross-entropy loss function predicts categorical values for each pixel.

3.4 SEGNET

SegNet model is an efficient encoder-decoder architecture explicitly designed for pixel-wise segmentation tasks [17]. The high-level feature representations from the input image are extracted by the encoder, while the decoder reconstructs these features to form the segmentation mask. SegNet differs from CNN architecture, using the transposed convolutional layers in the decoder [12].

The input size of SegNet model is 224 * 224 *1 as it uses grayscale. The encoder has three blocks of two layers comprised of 2D convolutional and max-pooling layers with increasing filter sizes. The filters in each convolutional layer are 3x3. ReLU activation a function is introduced for non-linearity after each convolution. The convolutional layer extracts features, and the pooling layers, with a stride of 2x2, reduces the resolution of the features thereby the algorithm can focus on the most critical high-level features by reducing the computational time.

The decoder uses transposed convolutional layers to upsample the feature maps. The feature maps are directly transferred from the encoder to the relevant decoder layer through the skip connection. In the decoder, the final layer is a 1x1 convolutional layer because of binary segmentation with a sigmoid activation function.

During the training process, model performance metrics, like accuracy, typically improves over each subsequent epoch. However, after a certain epoch, some of the metrics may decline due to overfitting, as the model begins to learn from noise and other irrelevant patterns in the training data. To avoid this, the training process should include a mechanism to evaluate the model after each epoch and terminate training when there is a decline in performance. This approach is known as early stopping [18]. In the TensorFlow library, the ReduceLROnPlateau function modifies the learning rate in response to the model's performance on a validation metric. This approach helps to avoid overfitting and promotes effective learning of the model. If the validation loss does not improve for 2 consecutive epochs, the learning rate is reduced by a factor of 0.5. This process continues until the learning rate reaches the minimum specified rate of 1e-6.

4. RESULTS

On executing the U-net model it achieved an accuracy of 0.8434, but the test accuracy was 0.8234, which shows that the model is overfitting. The training loss is lower than the validation loss (0.0411 vs. 0.3025), which shows that the model performs very well on the training data, but it struggled during the validation data set. Overfitting happens when the model learns the training data excessively well, capturing noise and irrelevant details, which ultimately results in poor performance on a new raw data. The computation time for DeepLabv3 model is very less when compared to other models used because of the pretrained weights. The training phase last for only 24 epochs with accuracy of 92%. The validation loss is calculated after each epoch during the training process using the validation set. The test loss is assessed after the completion of training process. The model is evaluated on a exclusive test data set, to assess the final model's performance in a real-world scenario. It was observed that the model is stable, as the difference between validation and test loss is very close.

The SegNet model achieved an accuracy of 0.8757 during training and 0.8745 and 0.8744 in validation and test phases at 156 epochs, indicating no overfitting and the well-trained model. The binary cross-entropy loss function provides a test loss value of 0.042, suggesting that the predicted probabilities are very close to the actual mask. Each model has its own feature selection process, which is one of the factors that determine its accuracy [3]. In segmentation task accuracy measures how closely, the models predicted segmentation mask matches the mask [14].

The accuracy [18] is computed based on the formula provided in Eq.(2)

$$Accuracy = \frac{(Number\ of\ correctly\ classified)}{(Total\ Images\ in\ the\ dataset)} \tag{2}$$

The spatial overlap between the original mask and the segmented image is measured by the Dice coefficient [22].

$$\text{Dice Coefficient} = \frac{2 \times \text{Intersection}}{\text{Size of Prediction} + \text{Size of Ground Truth}} \tag{3}$$

In Eq.(3), takes the intersection of the predicted and true regions and multiplies by 2 and divides by the sum of their sizes. Dice coefficient is more common in medical imaging, where small differences can be significant [19]. The training and validation of both accuracy and Dice coefficient over each epoch for SegNet model is plotted in Fig.1. During the initial epochs, there was a slight variation between training and validation, but after 80 epochs there was either marginal or nil deviation observed.

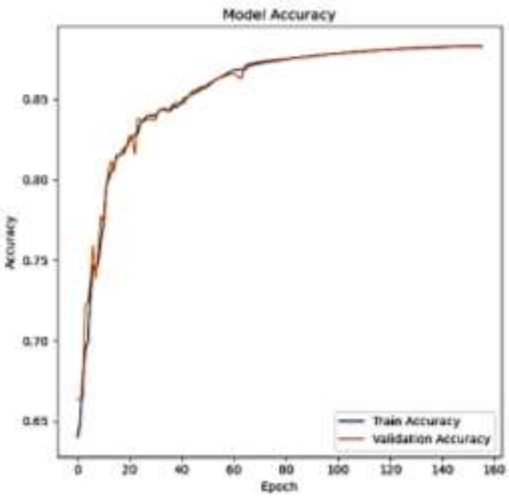

Fig.1. Plot of (a) Model training vs validation accuracy

The key performance metrics (accuracy, loss, and validation accuracy) is compared across three models (Table.1): SegNet, DeepLabv3, and U-Net. It is seen that in this study amongst the three models, DeepLabv3 has the highest accuracy of 0.9235, followed by U-net with 0.8901 and SegNet with 0.8757. The metric loss value has been analyzed. The loss value represents how well the model fits the data set after training. SegNet has the lowest loss value of 0.04, indicating better performance in terms of minimizing the error compared to DeepLabv3 with 0.18 and U-Net with 0.0468.

The validation accuracy metric compares how well the three model generalizes for unseen data. DeepLabv3 again shows the highest validation accuracy of 0.9278, followed by U-Net with 0.8886 and SegNet with the lowest validation accuracy of 0.8762. As per the above comparison of metrics, the DeepLabv3 is found to outperform the other two models in terms of both training and validation accuracy, even though SegNet achieves a low loss.

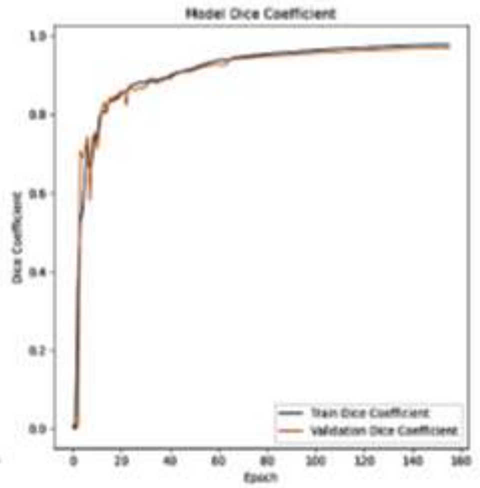

Fig.1. Plot of (b) Model training vs validation Dice coefficient, for each epoch

Table.1. Comparison of training accuracy, loss and validation accuracy for the SegNet, U-Net, and DeepLabv3 models

Metrics	SegNet	DeepLabv3	U-Net
Training Accuracy	0.8757	0.9235	0.8901
Loss	0.042	0.18	0.0468
Validation Accuracy	0.8762	0.9278	0.8886

The Mask prediction for this dataset by SegNet Model, using original image, ground truth mask and predicted mask is shown in Fig.2. The predicted mask is compared with ground truth mask to assess the model's performance in segmentation.

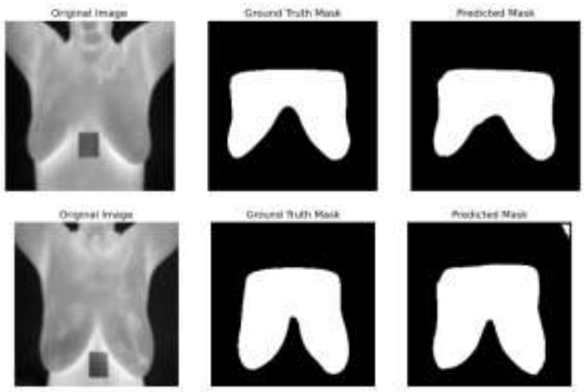

Fig.2. Original image with ground truth and predicted mask

The model trained in two different machines successfully, has been evaluated for performance based on computation time and by various evaluation metrics. The results for evaluation for all the three models are discussed.

4.1 PERFORMANCE ASSESSMENT COMPUTATIONAL TIME

As mentioned earlier all these models are executed on two different machines and the computation time taken by each model on Machine 1 and Machine 2 is shown in Table.2. It is observed that Machine 1 encountered kernel crash due to high computational demand.

Table.2. Computational Time for each model

Model	Machine 1	Machine 2
U-Net	≈12.6 hours	≈6.1hours
SegNet	≈41.3 hours	≈23.05 hours
DeepLabv3	≈6.2 hours	≈3.6 hours

As per the data in Table 2 SegNet is remarkably slower which may be attributed to operations like down sampling in a MaxPooling layer. The models, U-Net and SegNet are built from scratch, hence the training process is time-consuming and in such cases in order to optimize the computational time powerful GPUs may be used. The computation time for DeepLabv3 is remarkably less as it uses pretrained models for segmentation.

4.2 EVALUATION METRICS

Deep learning-based models can learn complex image features and produce precise segmentation results [20]. However, despite their accuracy, segmentation errors can still occur and are often categorized based on several factors: the quantity of segmented objects, the area of the segmented regions, the accuracy of boundary alignment, and the presence of internal holes or gaps in the boundaries of the segmented areas [21]. Metrics are used to capture and quantify the various types of errors depending on the data and the specific segmentation is listed.

4.2.1 Pixel Accuracy and Intersection of Union:

Pixel Accuracy measures how many individual pixels are correctly classified based on Eq.(4)

$$\text{Pixel Accuracy} = \frac{\text{No. of Correctly Classified Pixels}}{\text{Total Pixels}} \tag{4}$$

IoU, on the other hand, measures the overlap between the predicted and ground truth regions, specifically focusing on the foreground [16] by Eq.(5):

$$\text{IoU} = \frac{\text{Intersection of Predicted and Ground Truth}}{\text{Union of Predicted and Ground Truth}} \tag{5}$$

Evaluation by IoU involves identifying the incorrectly predicted and missed regions in ROI. Generally, models with high accuracy (ability to predict a large number of background pixels correctly), have a low IoU as it fails to correctly identify the boundaries of the foreground.

This study shows DeepLabv3 achieves the highest accuracy in both training and validation, showing its superior performance for segmentation tasks. Whereas its loss value (0.18) which is significantly higher compared to SegNet and U-Net. Lower IoU and Dice coefficient of DeepLabv3 indicates that it is difficult for it to accurately identify ROI and segmentation refinement.

SegNet model shows consistent training and validation accuracy, which implies good generalization. The low loss indicates that the model is confident in its predictions and shows balanced metrics, robust performance across IoU, Dice, Precision, and Recall. Also, the overlap (IoU) and similarity (Dice coefficient) for SegNet model is lesser than U-Net.

Though U-Net has the lowest accuracy among the three models both in training and validation, its loss is low indicating better confidence in predictions. The drop in validation accuracy for this model from 0.8901 to 0.8886 might suggest slight over fitting. But excluding accuracy other metrics shows that this model performs consistently well compared to the other two models.

Table.3. Comparison between SegNet, U-Net, DeepLabv3

Model	SegNet	U-Net	DeepLabv3
IoU	0.8726	0.9380	0.8057
Dice Coefficient	0.9307	0.9680	0.8924
F1 Score	0.9307	0.9680	0.8924
Precision	0.9217	0.9695	0.9291
Recall	0.9418	0.9664	0.8585
Pixel Accuracy	0.8744	0.8923	0.9273

The IoU metric is one of the most used to assess the overlap between the predicted and ground truth segmentation masks [22]. U-Net outperforms both SegNet and DeepLabv3 with an IoU of 0.9380, indicating that U-Net produces the most precise segmentation boundary.

Both the Dice Coefficient and F1 Score [23] reflect the balance between precision and recall, which are crucial for assessing the overall performance of segmentation models. The U-Net model leads with the highest Dice Coefficient and F1 Score of 0.9680, outperforming SegNet (0.9307) and DeepLabv3 (0.8924). This demonstrates that U-Net not only detects the relevant objects well but also minimizes false positives and false negatives, making it a more reliable model overall.

Precision is a measure of how many of the predicted positive pixels are correct [22]. U-Net achieved the highest precision score (0.9695), indicating it made fewer false positive predictions compared to SegNet (0.9217) and DeepLabv3 (0.9291). This suggests that U-Net is the most conservative model in terms of predicting positive pixels, ensuring that the predicted positive regions are more likely to be correct.

Recall measures how many of the actual positive pixels were correctly identified [22]. SegNet performed best in terms of recall (0.9418), indicating that it identified the largest proportion of true positive pixels. However, this comes at the cost of slightly lower precision, which infers that SegNet might produce more false positives compared to U-Net.

The Pixel Accuracy metric, which calculates the overall proportion of correctly classified pixels in the image, was highest for DeepLabv3 (0.9273), followed by U-Net (0.8923) and SegNet (0.8744). The DeepLabv3 showed the highest Pixel Accuracy. This result doesn't necessarily imply it performed best in terms of segmentation quality, as the IoU, Dice, and F1 Score are lower for this model, indicating that the pixel accuracy might be inflated by easier to classify the regions in the image.

5. SUMMARY

Deep learning-based models can learn complex image features and produce precise segmentation results. The performance of image segmentation is mainly dependent on the dataset used to train and evaluate the model. The algorithm that performs better for particular dataset may not give same performance for other dataset. In this article the segmentation of figshare dataset has been studied using the three models. The performance of the models is assessed based on studied various metrics. While comparing the performance of the models it is observed that the U-Net excels in IoU, Dice coefficient, and class-specific metrics (Precision, Recall). The computation time of the U-Net is much higher than other two models. SegNet shows strong performance, particularly in Recall, indicating SegNet is better at detecting the complete object, with slightly lower precision. DeepLabv3 performs well in Pixel Accuracy but lags behind in other segmentation metrics, indicating that DeepLabv3 may not be reliable for segmenting complex objects with high precision. Our Studies reveals that U-Net model might be appropriate for identifying ROI in breast thermal images. However, further work may involve optimizing DeepLabv3 to improve its segmentation quality.

REFERENCES

[1] G.S. Bhattacharyya, D.C. Doval, C.J. Desai, H. Chaturvedi, S. Sharma and S.P. Somashekhar, "Overview of Breast Cancer and Implications of Overtreatment of Early-Stage Breast Cancer: An Indian Perspective", *American Society of Clinical Oncology*, pp. 789-798, 2020.

[2] Medical Buyer, "India Has Alarming Ratio of One Radiologist per Lakh People", *Medical Buyer*, Available at: https://www.medicalbuyer.co.in/india-has-alarming-ratio-of-one-radiologist-per-lakh-people/, Accessed in 2024.

[3] S.S. Sudha, K.B. Jayanthi, C. Rajasekaran and T. Sunder, "Segmentation of ROI in Medical Images Using CNN-A Comparative Study", *Proceedings of International Conference on TENCON*, pp. 1-9, 2019.

[4] M. Kaura and V. Wasson, "ROI based Medical Image Compression for Telemedicine Application", *Procedia Computer Science*, Vol. 70, pp. 579-585, 2015.

[5] M.B. Rakhunde, S. Gotarkar and S.G. Choudhari, "Thermography as a Breast Cancer Screening Technique: A Review Article", *Journal of Clinical Medicine*, Vol. 12, No. 9, pp. 2785-2799, 2023.

[6] M.A.S. Ali, G.I. Sayed, T. Gaber, A.E. Hassanien, V. Snssel and L.F. Silva, "Detection of Breast Abnormalities of Thermograms based on a New Segmentation Method", *Proceedings of the Federated Conference on Computer Science and Information Systems*, Vol. 5, pp. 255-261, 2015.

[7] A. Hossam, H.M. Harb and H.M. Abd El Kader, "Automatic Image Segmentation Method for Breast Cancer Analysis using Thermography", *Journal of Engineering Sciences, Assiut University, Faculty of Engineering*, Vol. 46, No. 1, pp. 12-32, 2018.

[8] N. Scales, C. Herry and M. Frize, "Automated Image Segmentation for Breast Analysis Using Infrared Images", *Proceedings of International Conference of Engineering in Medicine and Biology Society*, pp. 1-10, 2004.

[9] A. Lou, S. Guan, N. Kamona and M. Loew, "Segmentation of Infrared Breast Images using MultiResUnet Neural Networks", *Applied Imagery Pattern Recognition Workshop*, pp. 1-8, 2020.

[10] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, Vol. 9341, pp. 234-241, 2015.

[11] A. Dalmia, S.T. Kakileti and G. Manjunath, "Exploring Deep Learning Networks for Tumour Segmentation in Infrared Images", *Quantitative InfraRed Thermography Journal*, Vol. 17, No. 3, pp. 153-168, 2018.

[12] P. Vianna, R. Farias and W.C.A. Pereira, "U-Net and SegNet Performances on Lesion Segmentation of Breast Ultrasonography Images", *Research on Biomedical Engineering*, Vol. 37, No. 2, pp. 171-179, 2021.

[13] S. Alqazzaz, X. Sun, X. Yang and L. Nokes, "Automated Brain Tumor Segmentation on Multi-Modal MR Image using SegNet", *Computational Visual Media*, Vol. 5, No. 2, pp. 209-219, 2019.

[14] L. Ahmed, M.M. Iqbal, H. Aldabbas, S. Khalid, Y. Saleem and S. Saeed, "Image Data Practices for Semantic Segmentation of Breast Cancer using Deep Neural Network", *Journal of Ambient Intelligence and Humanized Computing*, Vol. 14, pp. 15227-15243, 2023.

[15] L.C. Chen, G. Papandreou, F. Schroff and H. Adam, "Rethinking Atrous Convolution for Semantic Image Segmentation", *Computer Vision and Pattern Recognition*, pp. 1-14, 2017.

[16] C.Y. Hsu, R. Hu, Y. Xiang, X. Long and Z. Li, "Improving the Deeplabv3+ Model with Attention Mechanisms Applied to Eye Detection and Segmentation", *Mathematics*, Vol. 10, No. 15, pp. 1-20, 2022.

[17] V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 12, pp. 2481-2495, 2017.

[18] L. Prechelt, "*Early Stopping-but When?*", Neural Networks: Tricks of the Trade, pp. 53-67, 2000.

[19] K.H. Zou, S.K. Warfield, A. Bharatha, C.M.C. Tempany, M.R. Kaus, S.J. Haker, W.M. Wells III, F.A. Jolesz and R. Kikinis, "Statistical Validation of Image Segmentation Quality based on a Spatial Overlap Index", *Academic Radiology*, Vol. 11, No. 2, pp. 178-189, 2004.

[20] J. Ma, Y. He, F. Li, L. Han, C. You and B. Wang, "Segment Anything in Medical Images", *Nature Communications*, Vol. 15, No. 1, pp. 1-7, 2024.

[21] R. Shi, J. Ma, K.N. Ngan, J. Xiong and T. Qiao, "Objective Object Segmentation Visual Quality Evaluation: Quality Measure and Pooling Method", *ACM Transactions on Multimedia Computing, Communications and Applications*, Vol. 18, No. 3, pp. 1-19, 2022.

[22] G.V. Vlasceanu, N. Tarba, M.L. Voncila, C.A. Boiangiu and D.F. Teodor, "Selecting the Right Metric: A Detailed Study on Image Segmentation Evaluation", *Broad Research in Artificial Intelligence and Neuroscience*, Vol. 15, No. 4, pp. 295-318, 2024.

[23] A.A. Taha and A. Hanbury, "Metrics for Evaluating 3D Medical Image Segmentation: Analysis, Selection and Tool", *BMC Medical Imaging*, Vol. 10, No. 4, pp. 1-28, 2015.