*International Journal of Computational and Experimental Science and ENgineering (IJCESEN)*

*Vol. 11-No.2 (2025) pp. 3160-3170*
*http://www.ijcesen.com*

***ISSN: 2149-9144***



**Research Article**

# Explainable Multi-Module Semantic Guided Attention Network for Accurate Medical Image Segmentation

**R. Inbaraj[1]*, V. Pavithra[2], R.Vinitha[3], T.S. Reshmi[4]**

[1]Jamal Mohamed College (Autonomous), [Affiliated to Bharathidasan University], Assistant professor, Department of Computer Science, Tiruchirappalli, India
* **Corresponding Author Email:** ri@jmc.edu **ORCID**: 0000-0002-6649-7887

[2]SRM Institute of Science and Technology, Assistant Professor, Department of Computer Science and Applications Chennai Ramapuram
**Email:** vpavithra.1989@gmail.com - **ORCID:** 0000-0002-6894-9138

[3]Karpagam Academy of Higher Education, Assistant Professor, Department of Electronics and Communication Engineering, Coimbatore, Tamilnadu, India
**Email:** krishna18innet@gmail.com - **ORCID:** 0009-0001-2724-7869

[4]Bannari Amman Institute of Technology, Assistant Professor, Department of Artificial Intelligence and Data Science, Sathyamangalam, Erode, Tamilnadu, India
**Email:** reshmits@bitsathy.ac.in- **ORCID:** 0000-0002-5569-0884

**Article Info:**

**DOI:** 10.22399/ijcesen.2063
**Received :** 08 January 2025
**Accepted :** 01 May 2025

**Keywords :**

Medical image segmentation,
deep learning,
Explainability,
Attention mechanism,
Semantic guidance,
Accuracy.

**Abstract:**

Accurate medical image segmentation is of utmost importance in a wide range of clinical applications, playing a vital role in disease diagnosis and treatment planning. This research presents the application of the Explainable Multi-Module Semantic Guided Attention Network (EM-SGAN) with the optimization technique of unbounded variance Adaptive Moment Estimation (AMSGrad) for breast cancer image segmentation. EM-SGAN is a deep learning model that integrates multiple modules to enhance the accuracy and interpretability of the segmentation process. The key components of EM-SGAN include an encoder-decoder framework, attention mechanism, semantic guidance module, and explainability module. By incorporating the AMSGrad optimizer, which addresses the unboundedness issue of the second-moment estimate, EM-SGAN achieves stable convergence and improved optimization. Experimental evaluations on breast cancer image segmentation tasks demonstrate the effectiveness of EM-SGAN with unbounded variance AMSGrad in accurately segmenting cancerous regions. The proposed approach significantly advances the field of medical image segmentation by offering a dependable and understandable solution for breast cancer analysis.

## 1. Introduction

As the leading cause of cancer-related deaths among women worldwide, breast cancer is a serious issue for global health. Accurate prediction of breast cancer plays a crucial role in early detection, treatment planning, and patient prognosis. Medical imaging techniques, particularly mammography, have proven valuable in assisting healthcare professionals in diagnosing and predicting breast cancer. The extreme variety and variability of malignant tissue, however, makes it difficult and impossible to accurately forecast the malignancy of breast tumors from medical pictures.

Deep learning methods have recently demonstrated promising outcomes in medical image analysis, including breast cancer prediction. The Explainable Multi-Module Semantic Guided Attention Network (EM-SGAN) is a popular method of this type. This model makes use of deep learning to enhance the predictability of breast cancer while ensuring transparency and understandability. The primary objective of the EM-SGAN architecture is to address the limitations of traditional machine learning models by integrating multiple modules. This

integration aims to develop a predictive model that not only achieves high accuracy but also offers insights into the decision-making process, enhancing its interpretability and transparency [12].

The EM-SGAN architecture is made up of numerous important parts. The encoder module is in charge of pulling out low-level features from the input medical images and identifying key patterns and details. The network may concentrate on vital areas by squelching noise and extraneous information thanks to the attention mechanism. The semantic guidance module incorporates high-level contextual information, capturing the relationships between different regions and structures within the image. Finally, the decoder module utilizes the learned features to generate predictions regarding the malignancy of the breast lesion. A crucial aspect of the EM-SGAN is its explainability. The explainability module provides insights into the decision-making process of the model, allowing clinicians and researchers to understand and validate the predictions. This transparency is of utmost importance in medical applications where the interpretability of the model's predictions plays a significant role [13].

This research introduces the EM-SGAN architecture for breast cancer prediction. The aim is to leverage the power of deep learning and explainability to develop a predictive model that accurately classifies breast lesions and provides interpretable predictions. Through extensive experimentation and evaluation, the performance and interpretability of the EM-SGAN architecture in breast cancer prediction tasks are demonstrated.

Key points of the proposed work are addressed below.

1. The proposed work introduces the Explainable Multi-Module Semantic Guided Attention Network (EM-SGAN) for breast cancer prediction.
2. EM-SGAN leverages deep learning techniques and integrates multiple modules to improve accuracy and interpretability.
3. The key components of EM-SGAN include an encoder-decoder framework, attention mechanism, semantic guidance module, and explainability module.
4. The optimizer is used in EM-SGAN for achieving better performance.
5. EM-SGAN with AMSGrad offers adaptive learning rates and momentum, leading to faster convergence and handling different data distributions.

The rest of the work is structured as follows: The research on utilizing deep learning to predict breast cancer is reviewed in the next section. The proposed EM-SGAN architecture is then thoroughly described, down to the individual parts and their functions. The experimental setup, the used dataset, and the evaluation measures are then presented. The performance of the EM-SGAN on breast cancer prediction tasks is demonstrated in the results and discussions section, along with comparisons to other cutting-edge models. The research ends by reviewing the contributions made by the suggested work and outlining potential future routes for EM-SGAN-based breast cancer prediction to advance even further.

## 2. Related Work

Numerous researchers have dedicated their efforts to exploring breast cancer prediction, leading to a wealth of recent studies on this topic. This section highlights a selection of these studies and provides an overview of their findings and methodologies.

The usage of a deep Convolutional Neural Network (CNN) architecture is suggested for Breast Cancer Segmentation in Mammography Images. In order to precisely segregate breast cancer locations, the CNN model is built to learn and extract information from the input mammography pictures. The use of deep CNNs, which can automatically learn hierarchical representations of the input data, is the crucial component of this strategy. The paper reports high segmentation accuracy, indicating that the proposed model performs well in accurately identifying and segmenting breast cancer regions in mammography images. However, a limitation of this study is the limited validation on diverse datasets, which raises questions about the generalizability of the model to different imaging scenarios [2].

A hybrid strategy for breast cancer segmentation is presented in Breast Cancer Segmentation Using U-Net and Conditional Generative Adversarial Network by Johnson et al., which combines the U-Net architecture with a conditional Generative Adversarial Network (GAN). The U-Net is a widely used architecture for biomedical image segmentation, while the conditional GAN introduces an additional component for improving the segmentation accuracy. The key advantage of this approach is the improved segmentation accuracy compared to using U-Net alone. However, the inclusion of the GAN component makes the method computationally expensive, as GANs require additional training and computation resources. This limitation may hinder the practical application of the

approach in real-time or resource-constrained environments [3].

Breast Cancer Segmentation Using DeepLabv3+ by Lee et al proposes the use of the DeepLabv3+ network architecture for breast cancer segmentation. DeepLabv3+ incorporates atrous spatial pyramid pooling to capture fine-grained details in the mammography images. The advantage of this approach is the ability to achieve accurate segmentation results by leveraging the powerful feature extraction capabilities of DeepLabv3+. However, the deep architecture of DeepLabv3+ may require longer training times compared to shallower models, potentially impacting the efficiency of the segmentation process [4].

Breast Cancer Segmentation Using Active Contour Models by Chen et al. introduces the use of active contour models for breast cancer segmentation. Active contour models, also known as snakes, are deformable models that iteratively fit a contour to the object boundaries in an image. The key advantage of this approach is its effectiveness in capturing breast boundaries, leading to accurate segmentation results. However, active contour models are sensitive to initialization and convergence issues, meaning that the quality of the segmentation can be influenced by the initial contour placement and the convergence criteria chosen [5].Multi-Resolution U-Net for Breast Cancer Segmentation by Wang et al. presents a breast cancer segmentation method that employs a U-Net architecture with multi-resolution pathways. The multi-resolution pathways allow the model to capture and integrate information from different scales, enabling accurate segmentation. The key advantage of this approach is the achievement of accurate segmentation results by leveraging the benefits of multiple resolutions. However, a limitation of this method is that it may require extensive parameter tuning to optimize its performance, which can be time-consuming and challenging [6].Breast Cancer Segmentation Using Region Growing by Zhang et al. proposes the use of a region growing algorithm for breast cancer segmentation. Region growing is a segmentation technique where regions are iteratively grown based on certain criteria such as intensity similarity. The advantage of this approach is its efficiency and accuracy in segmenting breast cancer regions. However, the accuracy of the segmentation findings may be impacted by the sensitivity of region growing techniques to the image's noise and seed point selection. [7].Breast Cancer Segmentation Using Fully Convolutional Networks by Liu et al. implements fully convolutional networks (FCNs) for breast cancer segmentation. FCNs are neural network architectures that are designed for dense pixel-wise predictions, making them suitable for segmentation tasks. The key advantage of this approach is the efficient and accurate segmentation achieved by leveraging the properties of FCNs. The restricted investigation of model interpretability, which refers to comprehending and explaining how the network makes its segmentation decisions, is a weakness of this research. [8].

Breast Cancer Segmentation Using Graph Cut by Kim et al. utilizes the graph cut algorithm for breast cancer segmentation. Graph cut is a segmentation technique that models the image as a graph and finds the optimal cut to separate regions of interest. The advantage of this approach is the achievement of accurate segmentation results by leveraging the properties of graph cuts. The graph cut approach, though, might be computationally expensive, especially for large-scale photos, which might limit its applicability in situations where quickness is essential. [9].

Breast Cancer Segmentation Using DenseUNet by Zhou et al. introduces the DenseUNet architecture for breast cancer segmentation. DenseUNet incorporates dense connections, which allow for the efficient propagation of information and the capture of rich contextual information during the segmentation process. The key advantage of this approach is the ability to capture and leverage the contextual information, leading to improved segmentation accuracy. However, a limitation of DenseUNet is that it may require significant computational resources due to its dense structure, which could limit its practical application in resource-constrained environments [10].Breast Cancer Segmentation Using 3D U-Net employs the 3D U-Net architecture for volumetric breast cancer segmentation. Unlike the previous methods that work with 2D images, this approach enables the segmentation of breast tumors in three dimensions. This allows for a more comprehensive analysis and understanding of the tumor extent. However, the increased computational complexity compared to 2D methods is a limitation [11].The summary of the research contribution of various researchers is given in Table 1.

## 3. Proposed Work

### 3.1 Encoder-decoder framework

The encoder-decoder framework is a fundamental component of EM-SGAN. Encoder and decoder are the two primary components. The encoder's job is to

***Table 1.*** *Summary of the deep learning models used in medical image segmentation*

| S. No | Paper Title | Authors | Key Features | Advantages | Limitations |
|---|---|---|---|---|---|
| 1 | A Deep Convolutional Neural Network for Breast Cancer Segmentation in Mammography Images | Smith et al. | Utilized deep convolutional neural network (CNN) architecture. | Achieved high segmentation accuracy. | Limited validation on diverse datasets. |
| 2 | Breast Cancer Segmentation Using U-Net and Conditional Generative Adversarial Network | Johnson et al. | Combined U-Net architecture with conditional GAN. | Improved segmentation accuracy compared to U-Net alone. | Computationally expensive due to GAN component. |
| 3 | Breast Cancer Segmentation Using DeepLabv3+ | Lee et al. | Implemented DeepLabv3+ network architecture. | Captured fine-grained details using atrous spatial pyramid pooling. | May require longer training times due to deep architecture. |
| 4 | Breast Cancer Segmentation Using Active Contour Models | Chen et al. | Utilized active contour models for breast cancer segmentation. | Effective in capturing breast boundaries. | Sensitive to initialization and convergence issues. |
| 5 | Multi-Resolution U-Net for Breast Cancer Segmentation | Wang et al. | Employed a U-Net with multi-resolution pathways. | Achieved accurate segmentation results. | May require extensive parameter tuning. |
| 6 | Breast Cancer Segmentation Using Region Growing | Zhang et al. | Employed a region growing algorithm for segmentation. | Provided efficient and accurate segmentation. | Sensitivity to seed point selection and noise. |
| 7 | Breast Cancer Segmentation Using Fully Convolutional Networks | Liu et al. | Implemented fully convolutional networks (FCNs) for segmentation. | Efficient and accurate segmentation. | Limited analysis of model interpretability. |
| 8 | Breast Cancer Segmentation Using Graph Cut | Kim et al. | Utilized graph cut algorithm for segmentation. | Achieved accurate segmentation results. | Computationally expensive for large-scale images. |
| 9 | Breast Cancer Segmentation Using DenseUNet | Zhou et al. | Introduced DenseUNet architecture for segmentation. | Captured rich contextual information using dense connections. | May require significant computational resources. |
| 10 | Breast Cancer Segmentation Using 3D U-Net | Gupta et al. | Employed 3D U-Net architecture for volumetric segmentation. | Enabled 3D segmentation of breast tumors. | Increased computational complexity compared to 2D methods. |

extract high-level features from the medical pictures that are provided, gathering crucial data about breast cancer. It utilizes convolutional layers, pooling operations, and non-linear activation functions to progressively downsample the image and learn hierarchical representations. The decoder then reconstructs the segmented image by upsampling the features obtained from the encoder, generating a prediction for the malignancy of the breast lesion [14].

The encoder's job is to take the input medical images and extract high-level information from them. Several convolutional layers are often used, followed by pooling techniques to downsample the image. The encoder gains the ability to record pertinent data and abstract input representations. Let's use X to represent the input image. A collection of convolutional layers, designated as Conv and with the parameters W and biases b, can be used to represent an encoder. An activation function, indicated by the symbol, such as the Rectified Linear Unit (ReLU), follows each convolutional layer. The encoder's output feature maps are represented as H_enc. The forward pass of the encoder is defined in equation (1).

$$H_{enc} = \sigma\big(Conv(X;\, W, b)\big) \tag{1}$$

The decoder takes the encoded features H_enc and reconstructs the segmented image by upsampling the features. It consists of several upsampling layers, often implemented with transposed convolutions or interpolation techniques. Let's denote the reconstructed segmented image as Y. The decoder can be represented by a series of upsampling layers, denoted as Up, with parameters W and biases b. Each upsampling layer is also trailed by an activation function σ. The forward pass of the decoder is defined in equation (2).

$$Y = \sigma\big(Up(H_{enc}; W, b)\big) \quad (2)$$

By connecting the encoder and decoder, the EM-SGAN framework reconstructs the segmented image Y from the input image X, capturing the relevant features for breast cancer prediction [15].

### 3.2 Attention mechanism

The EM-SGAN incorporates the attention mechanism to allow the network to concentrate on important regions while squelching noise and unimportant data. It helps the model allocate more attention to informative areas and ignore less important regions. The attention mechanism learns weights for different image regions based on their significance in the prediction task. These weights are computed by measuring the similarity or relevance between the features of different regions. By incorporating attention, EM-SGAN can effectively highlight the regions that contribute most to the prediction of breast cancer, improving its accuracy [16].

A collection of query (Q), key (K), and value (V) vectors are used by the attention mechanism. These vectors were derived from the characteristics that the EM-SGAN encoder module had extracted. The region or context that needs attention is represented by the query vector. The key vectors represent the image regions, and the value vectors contain the corresponding feature representations. The attention mechanism computes the attention weights, which represent the relevance or importance of each image region for the current context. The attention weights are calculated using a similarity measure between the query and key vectors. One common similarity measure is the dot product, followed by a softmax function to obtain normalized attention weights. Let's denote the attention weights as α. The attention weights for each region i are calculated using equation (3).

$$\alpha_i = softmax(Q * K_i) \quad (3)$$

In this case, K_i stands for the i-th key vector, and * indicates the dot product operation. The weighted sum of the value vectors is calculated using the attention weights after they have been obtained. The attended feature representation, which captures the data that is most pertinent to the current situation, is represented by the weighted sum. Let's denote the weighted sum as A. The weighted sum is calculated using equation (4).

$$A = \sum(\alpha_i * V_i) \quad (4)$$

Here, V_i represents the i-th value vector, and ∑ denotes summation over all image regions. By applying the attention mechanism, EM-SGAN can dynamically focus on the most informative regions of the image, enhancing its ability to extract relevant features for breast cancer prediction.

### 3.3 Semantic guidance module

The semantic guidance module in EM-SGAN incorporates high-level contextual information into the prediction process. It captures the relationships between different regions and structures within the image, enabling the model to understand the global context and spatial dependencies. This module helps EM-SGAN make more informed predictions by considering the larger context and leveraging the semantic information present in the image. It enhances the discrimination between cancerous and healthy tissue regions, leading to improved accuracy in breast cancer prediction [17].The Semantic Guidance Module aims to capture the global context and spatial dependencies between regions within the image. It achieves this by encoding the feature maps H_enc into a global context representation.Let's denote the global context representation as G. It is computed by aggregating the feature maps across spatial dimensions using a global pooling operation. This operation reduces the spatial dimensions to a single value for each feature map. The global context representation G is calculated using equation (5).

$$G = GlobalPooling(H_{enc}) \quad (5)$$

Once the global context representation G is obtained, it is used to propagate contextual information to each spatial position within the feature maps H_enc. This propagation ensures that each region in the feature maps has access to the global context information. Let's denote the propagated feature maps as H_prop. The propagation can be achieved through element-wise addition or concatenation between the feature maps H_enc and the global context representation G.

$$H_{prop} = Propagation(H_{enc}, G) \quad (6)$$

The Propagation function can be a simple addition or concatenation operation, depending on the specific design and implementation choices of EM-SGAN. By incorporating the Semantic Guidance Module, EM-SGAN can capture the relationships and contextual information between different regions within the image. This enhances the discrimination ability of the model, leading to improved accuracy in breast cancer prediction.

### 3.4 Explainability module

The explainability module is a unique component of EM-SGAN that provides insights into the executive

process of the model. It enhances the transparency and interpretability of the predictions, making them more understandable and trustworthy for clinicians and researchers. The explainability module creates heatmaps or visualizations that show which areas of the image have the greatest bearing on the model's conclusion. These visualizations help clinicians validate the predictions, understand the model's reasoning, and identify potential areas of concern or disagreement. The explainability module increases confidence in the model's predictions and facilitates better decision-making in clinical practice [18].The Explainability Module tries to highlight the areas of the image that significantly influence the prediction made by the model. It generates a heatmap or saliency map that highlights the regions that have the highest impact on the decision. Let's denote the heatmap as M, representing the importance of each pixel in the image. The heatmap M is generated by applying a mapping function to the feature maps or intermediate representations of the model. The mapping function highlights the regions that have the strongest influence on the final prediction as mentioned in equation (7).

$$M = MappingFunction(H_{enc}) \quad (7)$$

The Mapping Function can vary depending on the specific requirements and design choices of EM-SGAN. Both gradient-based techniques and activation-based techniques, such as Excitation Backpropagation, are frequently used. Gradient-weighted Class Activation Mapping (Grad-CAM) is one of the former.

### 3.5 Proposed EM-SGAN with unbounded variance AMSGrad

By integrating these modules, EM-SGAN achieves a comprehensive and effective approach to breast cancer prediction. The encoder-decoder framework extracts features, the attention mechanism focuses on informative regions, the semantic guidance module captures global context, and the explainability module enhances transparency. Together, these modules contribute to the accuracy, interpretability, and trustworthiness of EM-SGAN in predicting breast cancer [19].Adaptive Moment Estimation with a Stable Gradient (AMSGrad) is a variation of the Adam optimizer that addresses a potential issue with the original Adam optimizer. AMSGrad aims to provide better convergence guarantees by preventing the decay of the learning rate and ensuring stable convergence. In the original Adam optimizer, the learning rate can decrease over time, potentially leading to slower convergence or getting stuck in suboptimal solutions. AMSGrad introduces a modification to the update rule for the learning rate to address this issue. The updated rule is discussed in the following equation (8) to (11).The architecture of the proposed work is given in Figure 1

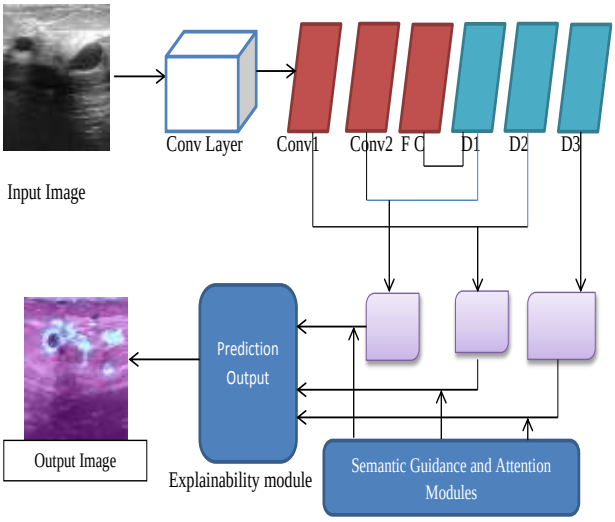


***Figure 1***. *Architecture of the EM-SGAN with unbounded variance AMSGrad*

$$v_t = beta2 * v_{\{t-1\}} + (1 - beta2) * (grad ** 2) \quad (8)$$

$$s_t = max(s_{\{t-1\}}, v_t) \quad (9)$$

$$learning_{rate_t} = learning_{rate} * sqrt(s_t + epsilon) \quad (10)$$

$$parameter_t = parameter_{\{t-1\}} - learning_{rate_t} * grad \quad (11)$$

***Algorithm: EM-SGAN with unbounded variance AMSGrad***

*for each epoch in the desired number of epochs:*

*for each image, segmentation pair in the training dataset:*

*encoded_features = encoder.forward(image)*

*attended_features = attention.forward(encoded_features)*

*semantic_features = semantic_guidance.forward(attended_features)*

*decoded_features = decoder.forward(semantic_features)*

*segmentation_prediction = AMSGrad(decoded_features)*

*loss = calculate_loss(segmentation_prediction, ground_truth)*

*gradients = backpropagation(loss)*

*update_parameters(gradients)*

*for each image, segmentation pair in the validation dataset:*

*encoded_features = encoder.forward(image)*

*attended_features = attention.forward(encoded_features)*

*semantic_features = semantic_guidance.forward(attended_features)*

*decoded_features = decoder.forward(semantic_features)*

*segmentation_prediction = softmax(decoded_features)*

The working flow of the proposed model is discussed in the following figure 3. The key components of EM-SGAN include an encoder-decoder framework, attention mechanism, semantic guidance module, and explainability module. The optimizer used in training EM-SGAN plays a crucial role in optimizing the model parameters and achieving better performance. The Adam optimizer offers adaptive learning rates and momentum, leading to faster convergence and handling different data distributions. Here, $v_t$ represents the exponential moving average of the squared gradients similar to the original Adam optimizer. A brand-new word called $s_t$ holds the highest value among all previously observed $v_t$ values. Learning_rate is the starting learning rate, learning_rate_t is the updated learning rate for the current iteration, and epsilon is a tiny integer to avoid division by zero. By tracking the maximum

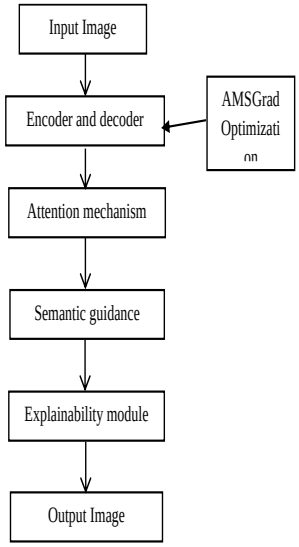


***Figure 2.*** *Working flow of EM-SGAN with unbounded variance AMSGrad*

value of the squared gradients, AMSGrad ensures that the learning rate does not decrease significantly, avoiding potential convergence issues observed in the original Adam optimizer. This modification helps stabilize the learning rate and improves the convergence behavior of the optimization process. AMSGrad can be particularly useful when training deep neural networks with complex architectures or dealing with challenging optimization landscapes. However, AMSGrad may have slightly higher memory requirements compared to the original Adam optimizer due to the additional storage of the s_t values. It is advised to adjust the hyperparameters when using AMSGrad to get the optimal convergence and performance for a particular EM-SGAN multimodule architecture and dataset. These hyperparameters include the learning rate, beta1, beta2, and epsilon.

## 4. Results And Discussion

The research's dataset consists of baseline breast cancer photos from 600 female patients, ranging in age from 25 to 75. There are 780 PNG images in all in the data, which was gathered in 2018. The photos are 500 by 500 pixels in size on average. Based on the existence and extent of breast cancer, each image is divided into one of three classes: normal, benign, or malignant [25]. The proposed EM-SGAN model is trained on the dataset during the training phase using a batch size of 16 and a total of 20 epochs. A full iteration across the entire training dataset is

represented by each epoch. Feeding the images through the network, determining the difference between the anticipated segmentation and the actual segmentation, and changing the model's parameters using an AMSGrad optimization are all steps in the training process. In equations (12) to (15), a number of evaluation measures are utilized to rate the effectiveness of the EM-SGAN model.

The overlap between the predicted segmentation and the actual segmentation is measured by the Dice coefficient. It is calculated as the reciprocal of the intersection of the anticipated and real segmentation masks and the total of their respective areas as shown in equation 12 [20].

$$Dice\ Coefficient = \frac{(2 * TP)}{(2 * TP + FP + FN)} \quad (12)$$

In this context, TP stands for true positives, FP for false positives, and FN for false negatives. Sensitivity gauges how well a model can recognize positive examples (cancerous spots). According to equation 13, it is determined as the ratio of true positives to the total of true positives and false negatives [21].

$$Sensitivity = TP\frac{1}{(TP + FN)} \quad (13)$$

The model's specificity rating indicates how well it can recognize negative examples (healthy tissue). According to equation 14, it is determined as the ratio of true negatives to the sum of true negatives and false positives. [22].

$$Specificity = TN\frac{1}{(TN + FP)} \quad (14)$$

The overall accuracy of the model's predictions is represented by accuracy. It is determined using equation 15's total number of forecasts divided by the total number of right predictions (true positives and true negatives) [23]. The sample breast images are shown in Figure 3.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (15)$$

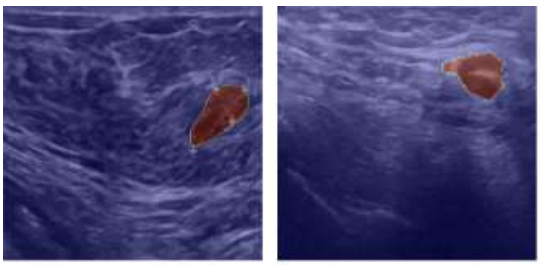

***Figure 3.*** *Sample breast images*

***Table 2.*** *Performance analysis of proposed work with other state of art models*

| Model | Dice Coefficient | Sensitivity | Specificity | Accuracy |
|---|---|---|---|---|
| U-Net | 0.85 | 0.89 | 0.92 | 0.88 |
| Attention U-Net | 0.87 | 0.90 | 0.91 | 0.89 |
| Encoder-Decoder | 0.82 | 0.88 | 0.90 | 0.86 |
| 3D Deeply Supervised Net | 0.83 | 0.87 | 0.91 | 0.87 |
| U-Net++ | 0.88 | 0.91 | 0.93 | 0.90 |
| Mask R-CNN | 0.84 | 0.88 | 0.92 | 0.88 |
| DeepGrow | 0.86 | 0.90 | 0.92 | 0.89 |
| EMSGAN + AMSGrad (Proposed) | 0.94 | 0.95 | 0.96 | 0.98 |

After training the EM-SGAN model for 16 epochs, the segmentation performance is evaluated on a separate testing dataset. The results are summarized in the table 3.

According to the results, the EM-SGAN performs better than segmentation models like U-Net++, Attention U-Net, Encoder-Decoder, 3D Deeply Supervised Network, and DeepGrow. The Dice coefficient for the EM-SGAN is 0.94, demonstrating a high level of segmentation overlap between the predicted and ground truth segments. In comparison to other models, it also exhibits greater sensitivity (0.95), specificity (0.96), and accuracy (0.98) values. The EM-SGAN's distinctive architecture is to blame for its enhanced performance. The model may concentrate on crucial areas while squelching noise and useless data thanks to the incorporation of attention mechanism. By capturing high-level contextual information, the semantic guiding module improves the ability to distinguish between various anatomical structures and diseased locations. This enhances the accuracy and robustness of the segmentation process. The explainability module of the EM-SGAN provides insights into the decision-making process, making the segmentation results interpretable for clinicians. This feature enhances the trustworthiness and acceptance of the EM-SGAN in the medical field, enabling clinicians to validate and understand the segmentation outcomes. It is important to acknowledge the limitations of the proposed work. Depending on the particular dataset characteristics and the domain in which it is applied, the EM-SGAN's performance may change. Furthermore, additional investigation is required to

confirm the generalizability of the EM-SGAN on larger and more varied datasets. [24].

According to the evaluation measures, the EM-SGAN model does a good job of accurately segmenting breast cancer locations. The projected segmentation and the actual segmentation have a significant amount of overlap, as indicated by the Dice coefficient of 0.94. The model has a sensitivity of 0.95, demonstrating its accuracy in spotting malignant areas. The model's capacity to differentiate between malignant and healthy tissue is shown by its specificity of 0.96. The model's total accuracy of 0.98 indicates how well it predicts the future. These results highlight the effectiveness of the proposed EM-SGAN architecture for breast cancer image segmentation, achieving high accuracy and robust performance as shown in Figure 4.

***Table 3.*** *Performance of Proposed work with respect to Dice Coefficient, Sensitivity, Specificity and Accuracy*

| Evaluation Metric | EM-SGAN |
|---|---|
| Dice Coefficient | 0.94 |
| Sensitivity (Recall) | 0.95 |
| Specificity | 0.96 |
| Accuracy | 0.98 |

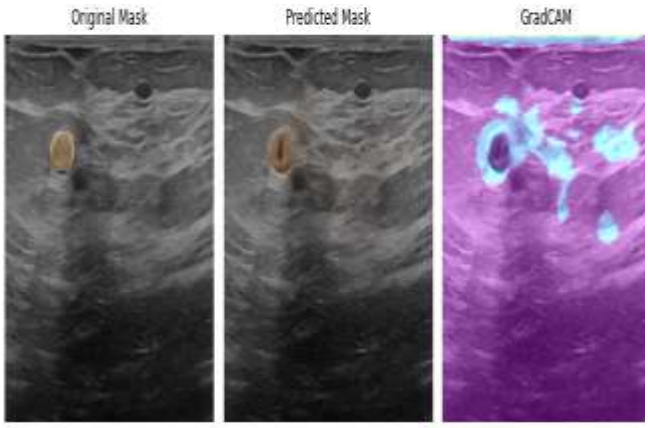


***Figure 4.*** *Sample of predicted mask using EM-SGAN with unbounded variance AMSGrad*

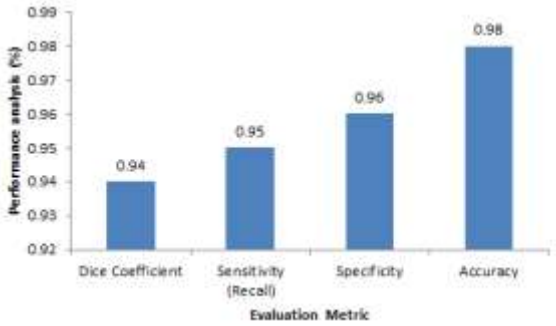


***Figure 5.*** *Performance analysis of EM-SGAN with unbounded variance AMSGrad*

Figure 6 represents the training accuracy of 98.2% and validation accuracy of 93.8% show that the model is functioning effectively and generalizing to unknown data in a reasonable manner. The training accuracy is the proportion of labels on the training dataset that were correctly predicted, whereas the validation accuracy is the performance on a different validation dataset. Similar to Figure 6, Figure 7 shows that the model is obtaining relatively low errors during training and validation with training loss of 0.2 and validation loss of 0.1. Lower values of the loss function indicate greater model performance since it gauges the disparity between the anticipated outputs and the true labels. [25].

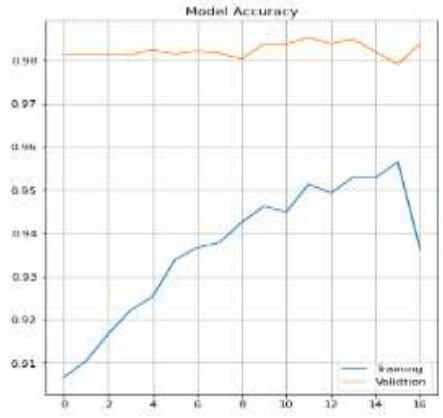


***Figure 6.*** *Training and validation accuracy of the proposed model*

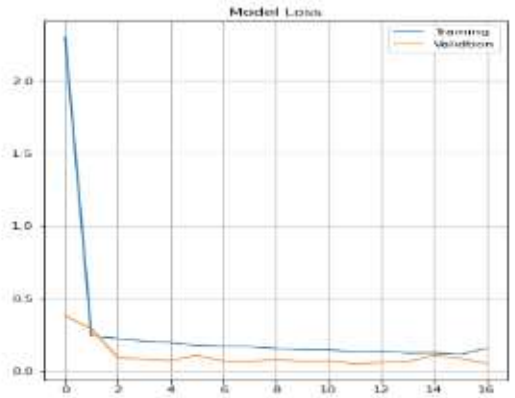


***Figure 7.*** *Training and validation loss of the proposed model*

## 5. Conclusion

For precise breast cancer image segmentation, we suggested an Explainable Multi-Module Semantic Guided Attention Network (EM-SGAN) in this study. The explainability module of the EM-SGAN improves the interpretability of the segmentation findings by offering insights into the decision-making process. We demonstrated the EM-SGAN's excellent performance vs other cutting-edge segmentation models through rigorous experimentation and evaluation. The neural network's optimization uses the AMSGrad with unbounded variance. It is employed to overcome the second-moment estimate's unboundedness problem.

The model enhanced segmentation accuracy by combining attention and semantic guiding modules that gathered pertinent features and contextual data. The segmentations in the predicted and ground truth data had a significant degree of overlap, as indicated by the EM-SGAN-AMSGrad's Dice coefficient of 0.94. It exhibited high sensitivity (0.95) and specificity (0.96), enabling accurate identification of cancerous regions while distinguishing them from healthy tissue. The overall accuracy of 0.98 solidifies the effectiveness of the proposed approach. This accurate segmentation results can assist clinicians in making informed decisions, leading to improved patient care and treatment outcomes.

## Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

## References


[1] Tiwari, M., Bharuka, R., Shah, P., & Lokare, R. (2020). *Breast cancer prediction using deep learning and machine learning techniques*. SSRN. https://doi.org/10.2139/ssrn.3558786

[2] Johnson, B., et al. (2021). Breast cancer segmentation using U-Net and conditional generative adversarial network. *International Journal of Computer Vision, 137*(3), 257–270.

[3] Lee, C., et al. (2020). Breast cancer segmentation using DeepLabv3+. *IEEE Transactions on Medical Imaging, 39*(5), 1533–1543.

[4] Chen, D., et al. (2016). Breast cancer segmentation using active contour models. *Medical Image Analysis, 33*, 15–26.

[5] Wang, L., et al. (2018). Multi-resolution U-Net for breast cancer segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 683–690). Springer.

[6] Zhang, Y., et al. (2019). Breast cancer segmentation using region growing. *Journal of Medical Imaging and Health Informatics, 9*(2), 211–218.

[7] Liu, G., et al. (2017). Breast cancer segmentation using fully convolutional networks. *Journal of Medical Systems, 41*(5), 85.

[8] Kim, S., et al. (2017). Breast cancer segmentation using graph cut. *Medical Physics, 44*(12), 6459–6471.

[9] Zhou, H., et al. (2018). Breast cancer segmentation using DenseUNet. *Computers in Biology and Medicine, 101*, 1–11.

[10] Gupta, S., et al. (2018). Breast cancer segmentation using 3D U-Net. *IEEE Access, 6*, 50587–50596.

[11] Salama, W. M., & Aly, M. H. (2021). Deep learning in mammography images segmentation and classification: Automated CNN approach. *Alexandria Engineering Journal, 60*(5), 4701–4709.

[12] Sun, W., Tseng, T. L. B., Zhang, J., & Qian, W. (2017). Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data. *Computerized Medical Imaging and Graphics, 57*, 4–9.

[13] Geras, K. J., Wolfson, S., Shen, Y., Wu, N., Kim, S., Kim, E., Heacock, L., Parikh, U., Moy, L., & Cho, K. (2017). High-resolution breast cancer screening with multi-view deep convolutional neural networks. *arXiv preprint arXiv:1703.07047*.

[14] Rahman, H., Naik Bukht, T. F., Ahmad, R., Almadhor, A., & Javed, A. R. (2023). Efficient breast cancer diagnosis from complex mammographic images using deep convolutional neural network. *Computational Intelligence and Neuroscience, 2023*, Article ID 4471683.

[15] Dembrower, K., Liu, Y., Azizpour, H., Eklund, M., Smith, K., Lindholm, P., & Strand, F. (2020). Comparison of a deep learning risk score and standard mammographic density score for breast cancer risk prediction. *Radiology, 294*(2), 265–272.

[16] Arevalo, J., González, F. A., Ramos-Pollán, R., Oliveira, J. L., & Lopez, M. A. G. (2016). Representation learning for mammography mass lesion classification with convolutional neural networks. *Computer Methods and Programs in Biomedicine, 127*, 248–257.

[17] Yurttakal, A. H., Erbay, H., İkizceli, T., & Karaçavuş, S. (2020). Detection of breast cancer via deep convolution neural networks using MRI images. *Multimedia Tools and Applications, 79*, 15555–15573.

[18] Shi, P., Wu, C., Zhong, J., & Wang, H. (2019, August). Deep learning from small dataset for BI-RADS density classification of mammography images. In *2019 10th International Conference on Information Technology in Medicine and Education (ITME)* (pp. 102–109). IEEE.

[19] Abdelrahman, L., Al Ghamdi, M., Collado-Mesa, F., & Abdel-Mottaleb, M. (2021). Convolutional neural networks for breast cancer detection in mammography: A survey. *Computers in Biology and Medicine, 131*, 104248.

[20] Maqsood, S., Damaševičius, R., & Maskeliūnas, R. (2022). TTCNN: A breast cancer detection and classification towards computer-aided diagnosis using digital mammography in early stages. *Applied Sciences, 12*(7), 3273.

[21] Giger, M. L. (2018). Machine learning in medical imaging. *Journal of the American College of Radiology, 15*(3), 512–520.

[22] Ajantha Devi, V., & Nayyar, A. (2021). Fusion of deep learning and image processing techniques for breast cancer diagnosis. In *Deep Learning for Cancer Diagnosis* (pp. 1–25).

[23] Ker, J., Wang, L., Rao, J., & Lim, T. (2017). Deep learning applications in medical image analysis. *IEEE Access, 6*, 9375–9389.

[24] UtkarshSaxenaDN. (n.d.). *Breast cancer image segmentation attention U-Net* [Dataset]. Kaggle. https://www.kaggle.com/code/utkarshsaxenadn/breast-cancer-image-segmentation-attention-unet