

Segmentation of Abnormalities in Mammogram Images for Early Breast Cancer Detection Using Mammo-Bench

Dr. Suba Suseela, Bonula Abhay Ankith, Shiv Shekhar

Department of Information Technology, Mahatma Gandhi Institute of Technology, Gandipet, Hyderabad, India

Department of Information Technology, Mahatma Gandhi Institute of Technology, Gandipet, Hyderabad, India

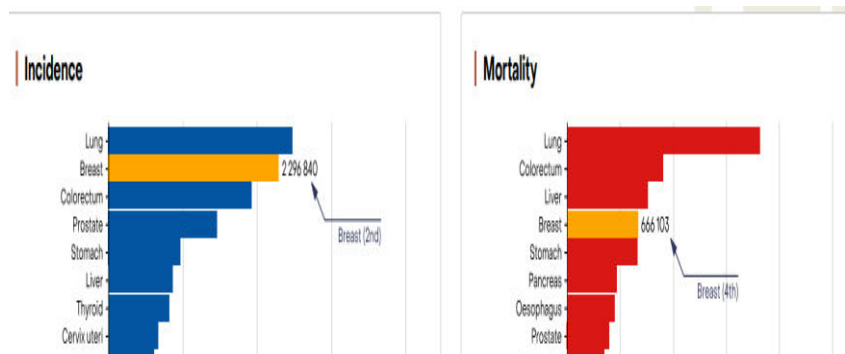
Department of Information Technology, Mahatma Gandhi Institute of Technology, Gandipet, Hyderabad, India

ABSTRACT: Breast cancer is one of the leading causes of mortality among women worldwide, and early detection plays a crucial role in improving survival rates. Mammography serves as the primary screening method which doctors use to detect breast abnormalities. The process of manually interpreting mammograms requires a long time and results in various human errors. This research presents an automated mammogram abnormality segmentation framework which uses the U-Net++ deep learning model for its operations. The system develops an end-to-end segmentation pipeline which precisely finds suspicious areas in mammographic images. The MammoBench multi-source dataset serves as the training data for the model, which combines multiple public mammography datasets including INBreast, Mini-DDSM, CDD-CESM, CMMD, DMID, KAU-BCMD, and RSNA Screening Mammography. The dataset requires preprocessing, which includes steps for normalization, format standardization, metadata unification, and ROI mask generation. The U-Net++ model achieves better segmentation results through its nested skip connections which enable better feature transmission and decrease the semantic connection gap between encoder and decoder components. The experimental results show that the proposed method successfully segments abnormal breast regions, which leads to better performance in computer-aided diagnosis systems

KEYWORDS Mammography, Breast Cancer Detection, Deep Learning, Medical Image Segmentation, U-Net++, MammoBench Dataset

I. INTRODUCTION

Breast cancer represents a major threat to world health, impacting numerous women across the globe. In reference to statistical analysis on cancer across the globe, it has been established that a considerable number of fatalities related to cancer occur within the female population due to breast cancer. Therefore, early diagnosis and correct diagnosis are extremely critical. The early diagnosis of medical conditions results in improved survival outcomes since it gives medical practitioners access to a number of alternatives for treating a patient. Mammography represents a prominent and reliable method for detecting abnormalities within the breasts through its various forms. Mammograms allow medical experts to identify minute structural changes within the breasts, which include but are not limited to, masses, calcifications, architectural distortions, and other areas indicating possible cancer development.



In 2022, there were about 2.3 million new cases of breast cancer diagnosed globally in women, and about 670,000 deaths from breast cancer. If current trends continue, by 2050 there may be 3.2 million new cases per year and 1.1 million deaths per year from breast cancer globally.

The interpretation of mammograms is found to be a difficult task for radiologists who have vast experience, as mammograms are composed of complicated tissue structures and varying densities, and low contrast between normal and abnormal tissues. The presence of dense breast tissues is found to be a significant challenge that acts as an obstacle in the detection of abnormalities, making the detection of abnormal tissues more difficult to achieve. The detection process involves the screening of massive mammograms, and it acts as an operational challenge that leads to diagnostic errors, such as false positives and undetected abnormalities. The detection of abnormalities should be achieved through automated computer-aided detection systems, as they solve the problems that are present in medical diagnostics.

Author & Year	Limitations / Issues	Existing Method	Proposed Method	Dataset Utilization	Techniques Applied	Tools Used	Validation Metrics	Future Directions
Bhole, Gaurav, S. Suba, and Nita Parekh (2025)	Class imbalance, demographic skew, and limited molecular subtype information in mammography datasets.	Previous research relied on fragmented and small mammography datasets.	Mammo-Bench integrates multiple datasets into a unified benchmark with standardized annotations.	Uses over 74,000 mammography images from about 26,500 patients across seven countries.	Data standardization, ROI annotation, dataset aggregation, and preprocessing.	Python, Deep Learning frameworks such as TensorFlow and PyTorch.	Dataset benchmarking metrics, classification accuracy, and evaluation consistency.	Improve dataset diversity, include more molecular subtype data, and enhance dataset balancing.
Joe Logan, Paul J. Kennedy, and Daniel Catchpole (2023)	Many mammography datasets are small, unbalanced, and lack standardized formats.	Traditional breast cancer research used independent datasets with limited accessibility and poor FAIR compliance.	Provides a systematic review of mammography datasets focusing on FAIR principles (Findable, Accessible, Interoperable, Reusable).	Examines multiple public breast cancer imaging datasets and evaluates their usability for research.	Dataset comparison, FAIR evaluation framework, metadata analysis.	Data management tools and research repositories.	Dataset accessibility metrics, reproducibility assessment, and FAIR compliance evaluation.	Encourage standardized dataset sharing and improve global access to breast imaging data.
Dajam, Doha Saad, and Ayman Qahmas h (2025)	Limited publicly available ultrasound datasets and reduced performance when applied to external	Conventional CNN-based classification models for breast tumor detection.	Deep learning framework combining classification and segmentation for breast cancer detection in	Uses benchmark ultrasound datasets for tumor detection and classification tasks.	Convolutional Neural Networks (CNN), image preprocessing, segmentation models.	Python, TensorFlow, Keras, OpenCV.	Accuracy, Precision, Recall, F1-score, Dice coefficient.	Improve domain adaptation and develop robust multi-task learning models for medical

Author & Year	Limitations / Issues	Existing Method	Proposed Method	Dataset Utilization	Techniques Applied	Tools Used	Validation Metrics	Future Directions
	datasets.		ultrasound images.					imaging.
Sarah M. Hooper et al. (2023)	Small labeled datasets cause overfitting and poor generalization in medical imaging models.	Fully supervised segmentation approaches that require large annotated datasets.	Semi-supervised segmentation methods that utilize both labeled and unlabeled medical images.	Uses multiple medical imaging datasets to test semi-supervised learning performance.	Semi-supervised learning, segmentation algorithms, data augmentation.	PyTorch, medical imaging toolkits.	Dice score, Intersection over Union (IoU), segmentation accuracy.	Develop more efficient semi-supervised techniques to reduce dependence on expert annotations.
Haochen Yuan and Junjie Peng (2024)	Low-contrast images cause unclear boundaries and noise leading to reduced segmentation performance.	Traditional semantic segmentation models struggle with low-contrast medical images.	LCSeg-Net model designed specifically for low-contrast image segmentation.	Evaluated on medical imaging datasets containing low-contrast lesions.	Deep learning segmentation network integrating structural and frequency information.	Python, PyTorch, deep learning frameworks.	Dice coefficient, IoU, segmentation accuracy.	Improve robustness of segmentation models and integrate attention mechanisms for better feature extraction.

Table 1: Literature Review and Dataset for Segmentation of Abnormalities in Mammogram Images for Early Breast Cancer Detection Using Mammo-Bench

II. LITERATURE REVIEW

Bhole, Suba, and Parekh [1] proposed a large-scale dataset for the purpose of research and computer-aided diagnosis systems for mammography images. The proposed dataset, known as Mammo-Bench, includes a unified dataset that combines seven publicly available datasets on mammography images. The dataset includes over 74,000 images obtained from more than 26,500 patients from seven different countries.

Dajam and Qahmash [2] developed a deep learning framework that classifies and segments the breast cancer tumor based on the ultrasound imaging technique. In this paper, the authors used the classification and segmentation techniques together to improve the performance of the deep learning model. In the experiment, the authors found that the proposed method showed high accuracy when tested with the standard dataset. The study by Logan, Kennedy, and Catchpoole [3] carried out an entire assessment of the mammography datasets to ensure that they conform to the FAIR principles of data management, which require that the datasets are findable, accessible, interoperable, and reusable. In this study, the research study focuses on various publicly available datasets for mammography, which are used by various researchers in the study of breast cancer.

In their research, Hooper et al. [4] investigated how effective semi-supervised learning approaches are in medical image segmentation problems. The process of developing fully annotated datasets for medical imaging studies requires significant time and financial investments, considering that expert radiologists are required to annotate each medical image. The research proves that semi-supervised learning approaches are effective in attaining performance results comparable to fully supervised approaches, considering that the research required fewer labeled images

Another model, LCseg-Net, was proposed by Yuan and Peng [5] and functions as a system for carrying out semantic segmentation. This model targets the segmentation of images, especially when the images bear little contrast between abnormal and normal areas.

III. PROPOSED METHODOLOGY

The proposed system utilizes a framework for segmentation, which is a form of deep learning. This framework is based on the U-Net++ architecture. U-Net++ acts as a form of improved framework for the U-Net framework, which medical professionals utilize for the segmentation of medical images. This framework includes the utilization of skip connections, which aid in the reduction of the semantic gap between complex and basic features.

The encoder extracts the hierarchical features from the inputted mammogram image, and the decoder performs the reconstruction of the segmentation map. This hierarchy of features enables the intermediate layers of the network to combine features from different levels, thus enhancing the segmentation results. This network produces a segmentation mask, showing the areas of abnormality present in the mammogram image.

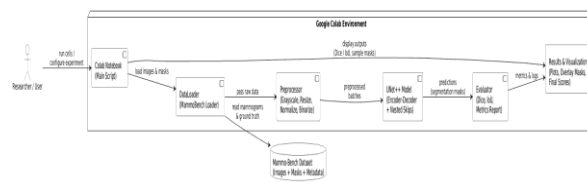


Fig 2: System Architecture

A. Data Acquisition and Preprocessing

The dataset used is the Mammo-Bench dataset, which contains standard mammogram images and their masks. The steps include converting the images to grayscale and resizing them to a particular size, 512x512, and normalizing their intensity, and converting the masks to binary and separating the lesions and their backgrounds.

B. Data splitting

The data is split into training, validation, and testing sets to ensure unbiased testing and prevent overfitting.

C. Augmentation techniques

Flipping, rotation, scaling, and contrast changes are used to make the model more robust and address the issue of variability in the dataset.

D. Model architecture (UNet++)

Nested paths with dense skip connections are used as encoder-decoder architectures to minimize semantic gaps.

E. Skip-connection refinement

Convolutional blocks are added on the skip paths to better align the encoded and decoded features.

F. Deep supervision strategy

Intermediate segmentation outputs are used to improve the training feedback.

G. Training process

The model is trained with a hybrid loss function, and an optimizer is used to ensure stable convergence and better lesion detection.

H. Evaluation metrics

Dice Coefficient and IoU are used to evaluate the segmentation performance.

I. Inference and visualization

The trained model outputs pixel-wise segmented maps highlighting masses, calcifications, distortions, and asymmetries, which can be compared against ground-truth masks.

IV. EXPERIMENTAL SETUP AND EVALUATION

The experimental setup describes the environment, dataset preparation, model configuration, and training procedures used to develop and evaluate the proposed mammogram abnormality segmentation system.

Dataset Description

The research utilized a dataset known as the Mammo-Bench dataset. The dataset acts as a standardized testing dataset for research conducted on mammography. The dataset, which is a unified dataset for mammography research, includes a collection of publicly available datasets for mammography. The dataset includes over 74,000 images from approximately 26,500 patients across seven countries. Each image includes important metadata related to the images. The metadata includes BI-RADS category breast density information and lesion information. The dataset includes ground truth segmentation masks that highlight the location of abnormal regions within the images. These abnormal regions include masses, calcifications, architectural distortions, and asymmetries.

Data Preprocessing

The mammogram images had to go through a series of preprocessing steps before the deep learning model could begin its training process. First, the images had to be converted to a grayscale format. This format would allow for a simpler representation and understanding of the structural composition. The images had to be resized to a standard size, 512 x 512 pixels.

Data Splitting

The data set was split into three sets to ensure the reliability of the results of the evaluation.

The Training Set served as the guide for the segmentation model.

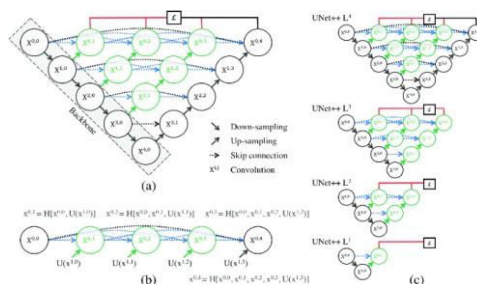
The Validation Set had two functions, and these included monitoring the progress of the training and the determination of the settings.

The Test Set served as the guide for the evaluation of the system's final performance. A standard ratio of 70% for the training needs and 15% for the validation and testing needs was used.

Data Augmentation

The medical imaging dataset had limited labeled data, which might result in overfitting. The researchers used various data augmentation techniques to produce more training data. This helped them to improve their existing training set. The techniques use three different methods which include:

- Random rotation
 - Horizontal and vertical flipping
 - Contrast adjustment
 - Random scaling and zooming
- Data augmentation enables the model to learn effective features. This will improve the model's ability to generalize on various types of mammogram images.



Model Configuration

The segmentation model used in this project is based on the architecture of the U-Net++ model. The U-Net++ is an enhanced version of the traditional U-Net model. The traditional U-Net++ model utilizes the nested encoder-decoder architecture of the U-Net++ model along with the dense skip connections to improve the propagation of the features and reduce the semantic gap between the encoder and the decoder of the network. The deep learning frameworks used were PyTorch and TensorFlow, along with Google Colab that offered the support of the GPU.

Key training parameters used in the experiment include:

- Optimizer:** Adam
- Loss Function:** Hybrid loss combining Dice Loss and Binary Cross Entropy (BCE)
- Learning Rate:** 0.001
- Batch Size:** 8–16 images per batch
- Number of Epochs:** 50–100 epochs depending on convergence

The hybrid loss function helps improve segmentation performance by balancing pixel-level accuracy and region-level overlap between predicted and ground-truth mask.

V. EVALUATION

The performance of the proposed segmentation model was evaluated using standard medical image segmentation metrics. These metrics measure how accurately the predicted segmentation mask matches the ground-truth annotation.

5.1 Evaluation Metrics

The following metrics were used to evaluate the segmentation performance:

Dice Coefficient

The Dice coefficient measures the overlap between the predicted segmentation and the ground truth mask.

$$L(Y, \hat{Y}) = -\frac{1}{N} \sum_{b=1}^N \left(\frac{1}{2} \cdot Y_b \cdot \log \hat{Y}_b + \frac{2 \cdot Y_b \cdot \hat{Y}_b}{Y_b + \hat{Y}_b} \right)$$

Y_b = ground truth segmentation

\hat{Y}_b = predicted probabilities

N = batch size

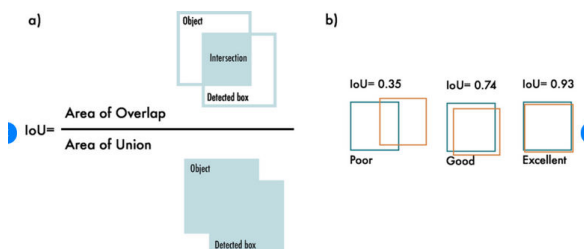
The second part of this equation:

$$\frac{2 \cdot Y_b \cdot \hat{Y}_b}{Y_b + \hat{Y}_b}$$

is the Dice Coefficient, which measures overlap between prediction and ground truth.

Intersection over Union (IoU)

The Intersection over Union metric measures the ratio of the overlapping region to the union of predicted and ground-truth masks.



Higher IoU values indicate better segmentation accuracy.

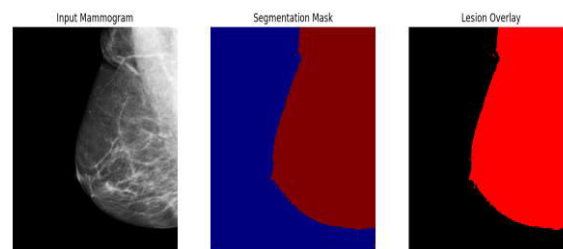
Precision and Recall

Precision measures the proportion of correctly predicted abnormal regions among all predicted abnormal regions, while recall measures the proportion of actual abnormal regions that were correctly detected. These metrics help evaluate the model's ability to detect lesions without producing excessive false positives.

5.2 Segmentation Results

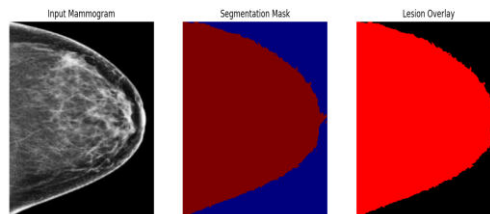
The capability of the U-Net++ model in segmenting abnormal regions in the mammographic images is shown in the experimental results. The U-Net++ model is able to segment various types of abnormal regions, such as masses and calcifications. The U-Net++ architecture is able to achieve accurate boundary detection using skip connections. The segmentation output of the U-Net++ model shows that the output is close to the ground truth segmentation provided in the Mammo-Bench dataset. The capability of the U-Net++ model in detecting suspicious regions in the mammogram images is shown in the results. The effectiveness of the proposed method is shown in the quantitative evaluation method used in the paper. The quantitative method uses the Dice and IoU metrics. The U-Net++ model is able to achieve good segmentation performance with stable results for various mammographic images.

VI. RESULTS



```

% AI CLINICAL REPORT
*****
Diagnosis: architectural distortion
Lesions: 5
Risk: High
Recommendation: Radiologist review + diagnostic imaging
Clinical_action: Urgent biopsy + oncological referral
*****
    
```



```

% AI CLINICAL REPORT
*****
Diagnosis: Mass Lesion
Lesions: 3
Risk: High
Recommendation: Radiologist review + diagnostic imaging
Clinical_action: Urgent biopsy + oncological referral
*****
    
```

```

...
Epoch 1/20 | Loss: 2405.0789 | Dice: 0.9973
✅ Phase-B best model saved
Epoch 2/20 | Loss: 1991.2574 | Dice: 0.9926
Epoch 3/20 | Loss: 1883.5371 | Dice: 0.9854
Epoch 4/20 | Loss: 1825.7647 | Dice: 0.9419
Epoch 5/20 | Loss: 1762.0787 | Dice: 0.9161
Epoch 6/20 | Loss: 1739.5600 | Dice: 0.9907
Epoch 7/20 | Loss: 1701.6692 | Dice: 0.7634
Epoch 8/20 | Loss: 1715.7195 | Dice: 0.9320
Epoch 9/20 | Loss: 1672.3909 | Dice: 0.8101
Epoch 10/20 | Loss: 1669.8376 | Dice: 0.9674
Epoch 11/20 | Loss: 1646.6190 | Dice: 0.9656
Epoch 12/20 | Loss: 1644.0712 | Dice: 0.9964
Epoch 13/20 | Loss: 1611.3262 | Dice: 0.9197
Epoch 14/20 | Loss: 1583.2848 | Dice: 0.9431
Epoch 15/20 | Loss: 1581.9121 | Dice: 0.9878
Epoch 16/20 | Loss: 1578.0844 | Dice: 0.9990
✅ Phase-B best model saved
Epoch 17/20 | Loss: 1550.7135 | Dice: 0.9970
Epoch 18/20 | Loss: 1557.4964 | Dice: 0.9849
Epoch 19/20 | Loss: 1558.6516 | Dice: 0.8804
Epoch 20/20 | Loss: 1550.4858 | Dice: 0.8183

✅ Phase-B Training Completed Successfully
    
```

VII. CONCLUSION

The project proposed a framework for deep learning, and the model utilized the U-Net++ architecture for the segmentation of abnormalities present in the images. This system detects abnormal areas present in the images of the mammary glands, and this helps doctors to identify breast cancer at the earliest stage. This model was trained and validated using the Mammo-Bench dataset, and the dataset includes a complete set of images of the mammary glands. The researcher utilized different preprocessing and data augmentation techniques to enable the model to handle different situations and to improve the model's performance.

The proposed model, U-Net++, effectively identifies abnormal areas present in the mammary glands, and these abnormal areas include masses and calcifications. The nested architecture of the model, U-Net++, along with its dense skip connections, enables the model to effectively detect abnormal areas present in the images of the mammary glands. Thus, the proposed system effectively demonstrates the capabilities of the proposed model for utilizing the capabilities of deep learning.

The proposed system effectively demonstrates the capabilities of the proposed model for utilizing the capabilities of deep learning.

VIII. FUTURE SCOPE

The suggested segmentation system shows promising results; however, in the future, there are various ways to improve this system. The system needs to classify the identified lesion as benign or malignant through the implementation of classification models. The implementation of classification models will enable a complete computer-aided diagnosis system.

The system currently performs better through the use of attention-based deep learning, which enables the system to identify the key areas of the images, thus enhancing the system's ability to segment the images. The research team also intends to improve the dataset by incorporating more images of mammography. This they plan to achieve by sourcing the images from various hospitals using different types of medical images. The model requires the improvement of the dataset since it enables the testing of the model's ability to perform well in different medical settings. The proposed system enables the integration of the system into the radiology of various clinical settings, which, coupled with the use of the PACS system, enables the provision of assistance to radiologists carrying out breast cancer screenings.

IX. REFERENCES

- [1] Bhole, Gaurav, S. Suba, and Nita Parekh. "Mammo-Bench: A Large-scale Benchmark Dataset of Mammography Images." medRxiv, 2025.
- [2] Dajam, Doha Sad, ad Ayman Qahmash. "Breast Cancer Classification and Segmentation Using Deep Learning on Ultrasound Images." International Journal of Advanced Computer Science & Applications 16.5 (2025).
- [3] Logan, Joe, Paul J. Kennedy, and Daniel Catchpoole. "A review of the machine learning datasets in mammography, their adherence to the FAIR principles and the outlook for the future." Scientific Data 10.1 (2023): 595.
- [4] Hooper, Sarah M., et al. "Evaluating semi-supervision methods for medical image segmentation: applications in cardiac magnetic resonance imaging." Journal of Medical Imaging 10.2 (2023): 024007-024007
- [5] Yuan, Haochen, and Junjie Peng. "LCSeg-Net: A low-contrast images semantic segmentation model with structural and frequency spectrum information." Pattern Recognition 151 (2024): 110428.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details