Attention-Guided U-Net with Gan Discriminator and CNN-Based Shape Classifier for Breast Tumor Detection

Kota Satyavathi Kumari[1], Kota Rahul Dev[2], Maharshi Dora[3]
[1,2,3] Student
Department of Computer Science and Engineering,
R.V.R. & J.C. College of Engineering, Chowdavaram, Guntur, Andhra Pradesh, India

ABSTRACT

Breast cancer causes the highest number of cancer- related deaths among women worldwide, making accurate classification and analysis of breast tumors in ultrasound imaging crucial for early diagnosis and treatment planning. In this work, we propose using a two-stage deep learning framework in segmentation and classification of breast ultrasound tumors analysis. In Stage I, we implement an Attention U- Net as a generator in a PatchGAN adversarial framework to perform segmentation of the tumor as accurately as possible. The model in Stage I is developed with adversarial loss and Tversky loss that allows us to overcome the class imbalance, and improve accuracy on the boundaries of the object. In Stage II, the tumor masks from Stage I are resized and classified as normal,benign,malignant using a conventional neural network (TumorNet).The masks are also classified as an irregular, lobular, oval and round tumors using a custom convolutional neural network (DeepShapeNet). The proposed framework was trained and validated on the Breast Ultrasound Images (BUSI) public dataset. The results demonstrate the segmentation model significant Dice scores and binary accuracy, and the classification model showed a strong ability to simply observe the shape and classify correctly based confusion matrix. As an integrated approach, the proposed framework offers a reliable and interpretable pipeline for providing automated assessment of breast tumors in clinical ultrasound imaging.

1

Keywords:

Breast cancer, Ultrasound imaging, Tumor segmentation, Attention U-Net, PatchGAN, CNN, Tumor shape classification, Tversky loss, Adversarial learning, Medical image analysis

INTRODUCTION

Breast cancer is a critical public health concern and the most commonly diagnosed cancer in women globally. It accounts for approximately 2.3 million cases annually, with mortality often resulting from delayed diagnosis and inadequate prognostic evaluation. Early detection of malignant tumors is essential for improving survival outcomes and guiding effective clinical intervention. Among various imaging modalities, ultrasound has emerged as a preferred diagnostic tool, particularly in developing countries and among younger population. Its non-ionizing nature, cost - effectiveness, and real-time imaging capability make it an indispensable component of routine breast cancer screening.

Despite its widespread use, ultrasound-based diagnosis presents unique challenges. Ultrasound images are often marred by speckle noise, artifacts, and low contrast between tumor and surrounding tissues. Moreover, tumors exhibit considerable heterogeneity in shape, size, and appearance. Manual interpretation by radiologists is time-intensive and prone to inter-observer variability. Misinterpretations can lead to false negatives or unnecessary biopsies, directly affecting patient outcomes. Consequently, there is a pressing need for robust, automated methods that can assist clinicians in segmenting tumor regions and characterizing their morphology with high precision.

Advancements in deep learning, particularly convolutional neural networks (CNNs), have revolutionized medical image analysis. Architectures like U-Net and its variants have demonstrated remarkable efficacy in segmentation tasks across modalities such as CT, MRI, and ultrasound. However, segmentation alone does not provide a comprehensive understanding of tumor behavior. In clinical settings, tumor shape is an important morphological feature used to infer the likelihood of malignancy. For instance, irregular and lobular shapes are often associated with malignancies, whereas oval and round shapes are more likely benign. Yet, traditional methods rarely integrate segmentation with morphological classification into a cohesive pipeline.

In this research, we propose a novel two-stage deep learning pipeline designed for the automatic segmentation and morphological classification of breast tumors in ultrasound images.

The pipeline consists of:

Stage I – Tumor Segmentation:

We use an Attention U-Net architecture that introduces attention gates to the conventional U - Net framework. This enables the network to focus selectively on tumor regions while minimizing the influence of irrelevant background structures. Attention mechanisms enhance spatial localization, which is critical for accurate tumor boundary delineation, especially in noisy ultrasound environments

Stage II – Tumor And Shape Classification:

Following segmentation, We introduce **TumorNet**, a CNN classifier designed to classify the segmented images into 3 predefined classes:

- Normal
- Benign
- malignant

We introduced DeepShapeNet, a lightweight CNN classifier specifically designed to categorize the binary tumor masks into four predefined morphological classes: irregular, lobular, oval, and round.

Both the classifiers are optimized with dropout and batch normalization for generalization, and uses large receptive field convolution layers to capture fine-grained shape variations.To evaluate our method, we employ two publicly available datasets:

- The **Breast Ultrasound Images Dataset** (BUSI), which includes benign, malignant, and normal cases along with expert-annotated masks.
- The **CBIS-DDSM** segmentation subset, known for high-quality annotated mammographic images that complement ultrasound-based features.

Comprehensive experiments were conducted to assess segmentation accuracy, classification precision, and generalization across both training and validation sets. Metrics such as confusion matrices, categorical accuracy, and training loss trends confirm the pipeline's effectiveness. Additionally, qualitative visualizations are presented to demonstrate the interpretability of both stages, with predicted masks and shape labels overlaid on original images.

This research aims to bridge the gap between image segmentation and high-level tumor interpretation, thereby enabling more informed diagnostic decisions. The proposed system not only supports radiologists in detecting tumors but also augments their decision-making with morphological insights that correlate with clinical malignancy patterns.

RELATED WORK

Automated analysis of breast tumors using deep learning has witnessed significant advancements in recent years, primarily focused on three interconnected tasks: lesion segmentation, tumor classification, and morphological analysis.

This section surveys the most pertinent developments in these areas, with special emphasis on works utilizing ultrasound imaging and shape-based diagnostic frameworks.

Breast Tumor Segmentation

Segmentation of breast tumors from ultrasound images has been extensively studied, given its importance in isolating the region of interest for further analysis. U-Net, introduced by Ronneberger et al., remains the foundational architecture for biomedical image segmentation, thanks to its encoder-decoder structure with skip connections that preserves spatial details. Numerous adaptations have since emerged. Oktay et al. proposed Attention U-Net, which integrates attention gates to enhance focus on salient regions, particularly beneficial in noisy environments like ultrasound scans. Several studies, including recent works by Al-Dhabyani et al., demonstrated the superior accuracy of attention-based segmentation over standard U-Net on BUSI and similar datasets.

Moreover, attempts have been made to combine segmentation with pre-processing enhancements such as speckle noise reduction or contrast enhancement. However, many of these approaches rely on handcrafted filters or pre-segmentation assumptions, limiting adaptability across imaging conditions. Our method bypasses such preprocessing dependencies by using a fully end-to-end trainable architecture with attention mechanisms embedded directly in the segmentation model.

Tumor Label Classification

Accurate classification of tumor types (e.g., benign, malignant) is critical for effective diagnosis and treatment planning. Traditional CADx systems relied on handcrafted features—such as texture, intensity, and edge descriptors—fed into machine learning classifiers like SVMs or Random Forests. However, these approaches are limited by their dependence on manual feature engineering and their sensitivity to image variability.

With the advent of deep learning, Convolutional Neural Networks (CNNs) have become the preferred approach for tumor classification, enabling direct learning from imaging data. Prior works, such as Yap et al., demonstrated the efficacy of deep CNNs in classifying breast lesions from full mammograms and ultrasound images.

In this study, we propose TumorNet, a CNN tailored for tumor label classification based on segmented tumor regions. Unlike methods that rely on entire images, our approach focuses on the tumor area, allowing the network to learn discriminative features from both morphology and intensity patterns, while reducing background noise. This region-specific design improves robustness and aligns with radiologists' clinical interpretation practices.

Tumor Shape Classification

Tumor shape provides crucial clinical information, often used by radiologists to distinguish between benign and malignant masses. While numerous CADx systems have been developed to classify tumor *types* (e.g., benign vs. malignant), relatively fewer studies have focused on *shape classification*. Traditional methods often use handcrafted features— such as compactness, roundness, or edge irregularity—extracted from binary masks and fed into classical machine learning models like SVM or Random Forests.

With the rise of deep learning, several researchers have adopted CNNs for direct classification of tumors. For example, Yap et al. developed a deep CNN-based framework for classifying breast lesions from mammograms and ultrasound images. However, most of these models operate on full images and do not explicitly isolate tumor shape from intensity and texture cues. In contrast, our approach uses segmentation-driven shape classification, where a CNN classifies the morphology of the tumor based purely on its binary mask. This separation ensures that shape, rather than intensity variance, drives the prediction.

Joint Segmentation and Interpretation

Only a limited number of studies have attempted to build pipelines that combine segmentation with downstream interpretative tasks such as classification. Liu et al. proposed a dual-stage framework where tumors are first segmented and then classified using separate models, but their approach focused on pathology types and not shape. Other works, such as those by Huang et al., integrate lesion characterization into end-to-end models using multi-task learning, yet these often blur the interpretability of intermediate results such as masks.

Our work distinguishes itself by offering a modular two-stage pipeline: the first stage precisely segments the tumor using Attention U- Net, and the second stage predicts the tumor type class and tumor shape class using a specialized CNN. This design not only promotes transparency but also enables each component to be optimized independently, ensuring better generalization and robustness.

Summary of Gaps

In summary, the following gaps exist in the current literature: A lack of pipelines that integrate accurate segmentation with clinically meaningful morphological classification.

- Limited use of attention mechanisms in segmentation models trained specifically on breast ultrasound datasets. Scarcity of deep learning approaches focusing exclusively on tumor shape classification from binary masks.
- Our research addresses these gaps by introducing a robust two-stage system combining attention-based segmentation with targeted morphological classification, validated on high-quality, expert- annotated datasets.

PROPOSED METHODOLOGY

Our system introduces a robust, two-stage pipeline for breast tumor shape classification directly from ultrasound images, combining the strengths of segmentation-based localization and mask-based classification. The approach is grounded in the principle of isolating shape information from visual noise, making it especially suited for low- contrast, artifact-prone modalities like ultrasound.

System Pipeline Overview

The end-to-end pipeline comprises two cascaded neural network stages:

Stage I: A customized Attention U-Net segments the tumor from the background.

Stage II: A handcrafted CNN classifies the tumor type and shape of the predicted tumor mask respectively into

- one of three predefined classes:normal,benign,malignant
- one of four predefined classes: irregular, lobular, oval, and round.

This modularity allows for separate optimization of both localization and classification, with interpretability preserved at each step. Figure 1 illustrates the complete flow.

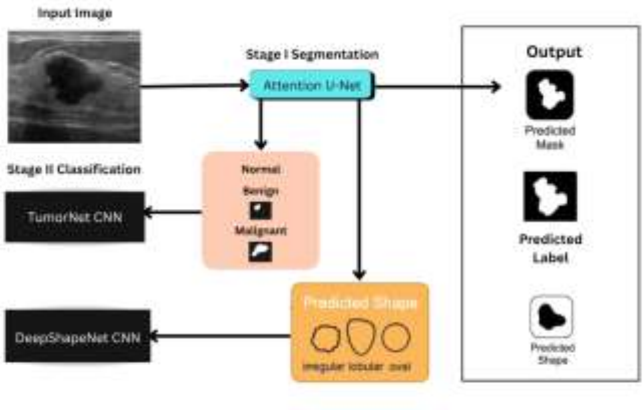

Fig 1. System Architecture

Stage I – Segmentation with Attention U-Net

Ultrasound images present challenges such as low contrast, noise, and artifacts. To handle these, we employ an Attention U-Net that introduces attention gates into the standard U-Net architecture. These gates learn to suppress irrelevant regions and enhance salient features, particularly around lesion boundaries.

Architecture Details:

Encoder Path: 4 levels of convolution + pooling (feature maps from 32 → 256).

Bottleneck: 512 filters capturing high-level features.

Decoder Path: Transposed convolutions with attention-guided skip connections.

Attention Mechanism: Each decoder level receives spatial attention from corresponding encoder maps, using gated feature fusion.

Output: Final sigmoid layer yields a probability map of tumor regions.

Training Objective:

To improve robustness against class imbalance, we use Tversky Loss, a generalization of Dice Loss, favoring a balance between sensitivity and specificity.

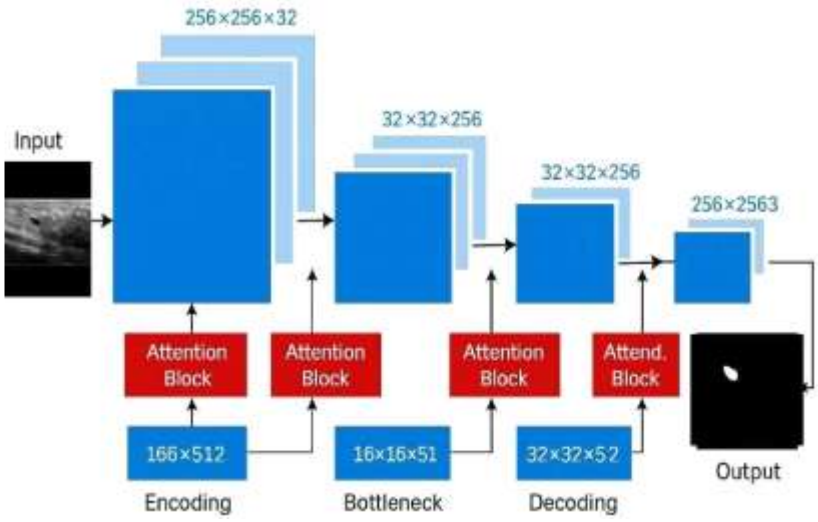

Fig 2. Generator network (Attention U-net) architecture

Post-Segmentation Processing

After segmentation:

- The output probability mask is thresholded at 0.5 to obtain a binary tumor mask.
- This binary mask is then resized to 64×64, removing scale variance and standardizing input for classification.
- Pixel intensity is not used; only the spatial geometry of the tumor is preserved, allowing the classifier to learn purely shape-based features.

This separation of concerns ensures that the CNN model does not rely on texture or brightness, which may vary across ultrasound machines or patients.

Stage 2 – Tumor Classification with TumorNet And Shape Classification with DeepShapeNet

A lightweight yet expressive CNN tailored for tumor and shape classification It focuses on structural morphology, not raw image content. TumorNet is designed to classify tumors into three categories using the segmented Image in the Stage 1 as input**.** This stage takes the processed binary mask as input. We introduce DeepShapeNet for shape recognition.

DeepShapeNet and TumorNet Architecture:

- Conv Layer 1: 64 filters, kernel size 9 → captures global shape features.
- Conv Layer 2: 128 filters, kernel size 5 → focuses on contours and edges.
- Conv Layer 3: 256 filters, kernel size 4 → deeper features.
- Fully Connected Head: Flattened vector → Dense(128) → Dense(4 softmax).

Design Considerations:

- MaxPooling (4×4) aggressively reduces dimensionality, retaining only shape outlines.
- Dropout (0.3 and 0.2) combats overfitting due to small input dimensionality.
- No texture-based filters used—only morphological features drive classification.

VOLUME: 09 ISSUE: 06 | JUNE - 2025 SJIF RATING: 8.586 ISSN: 2582-3930

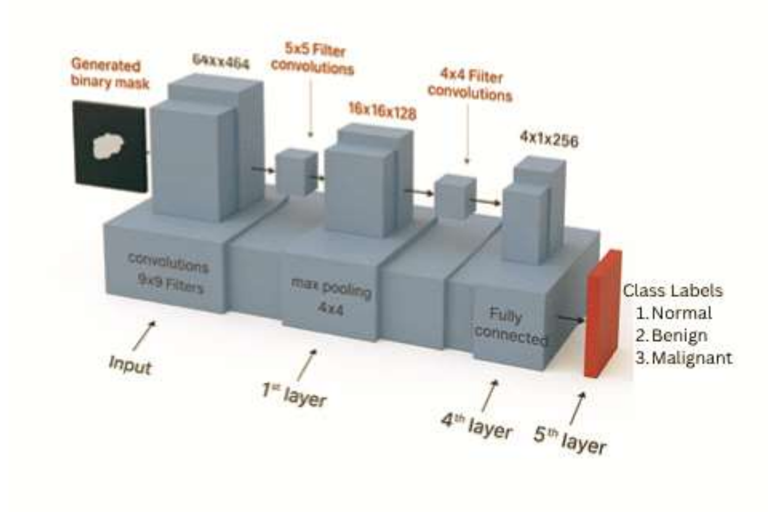

Fig 3.TumorNet Model Architecture

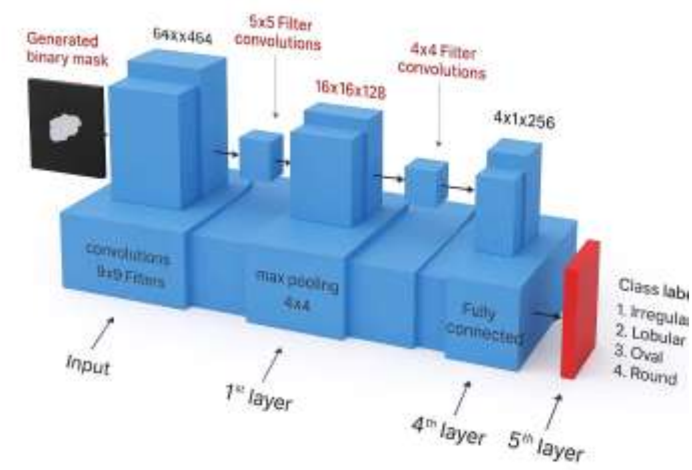

Fig 3.DeepShapeNet Model Architecture

Integration and End-to-End Inference

At inference time:

- The original image is passed through the trained Attention U-Net.
- The resulting tumor mask is thresholded and resized.
- The processed mask is input to TumorNet and DeepShapeNet.
- The model outputs tumor class label and shape class label with

associated confidence.

This pipeline is interpretable, reproducible, and modular. Additionally, mask predictions and class probabilities can be visualized for explainability.

RESULTS

- **Dataset and Preprocessing**

The experiments were conducted using the Breast Ultrasound Images (BUSI) dataset, which consists of 780 grayscale ultrasound images accompanied by ground truth binary masks indicating tumor regions. For our shape classification task, only the tumor-present images were selected and further annotated manually into four shape categories: irregular, lobular, oval, and round. This additional labeling facilitated supervised learning in Stage II.

Fig 4. Ultrasound Images

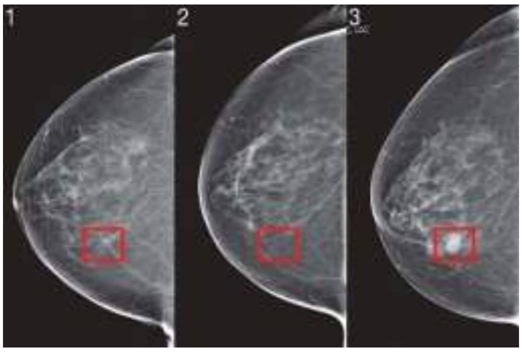

Fig 5. Curated Breast Imaging Subset of DDSM

All ultrasound images were resized to 256×256 pixels before being passed into the Attention U-Net for segmentation. The output binary masks were thresholded at 0.5 to obtain clear tumor boundaries. These masks were further resized to 64×64 pixels and normalized to serve as input to the CNN classifier. Image preprocessing steps included grayscale normalization and data augmentation techniques such as horizontal/vertical flipping, random rotations, and morphological enhancements (dilation/erosion), which helped in improving generalization.

Stage 1: Segmentation with Attention U-Net

The segmentation model was trained using the Attention U-Net architecture, which enhances the classical U-Net by integrating attention gates in the decoder path. These gates help the model focus on tumor-relevant features by suppressing irrelevant background regions. The model was optimized using the Tversky loss function, which is well-suited for imbalanced pixel distributions between foreground (tumor) and background.

Training was performed using the Adam optimizer with a learning rate of 1e-4 and batch size of 8. The model converged steadily with early stopping and learning rate reduction strategies (ReduceLROnPlateau) applied to avoid overfitting. Visual overlays between predicted and ground truth masks demonstrated strong boundary alignment, and the segmentation results achieved high Tversky scores and smooth validation curves.

Stage 2.1: Tumor Classification using TumorNet

The tumor classification model,TumorNet,was a custom Convolutional Neural Network that took segmentation mask images as input and predicts one of the three label categories.The network featured three convolutional blocks with dropout, batch normalization, and ReLU activations, followed by two fully connected layers for classification

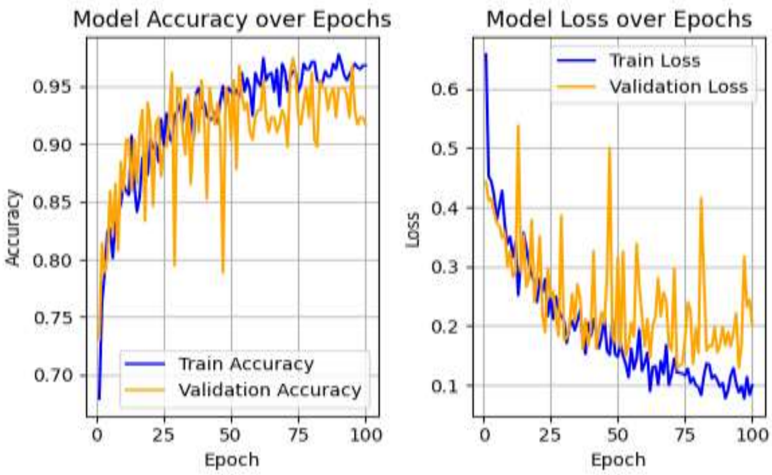

Fig6. Accuracy Graphs of TumorNet

The model was trained using categorical cross-entropy loss and the Adam optimizer with a learning rate of 1e-3. Early stopping was used based on validation accuracy. The training achieved 99% accuracy, while the validation accuracy reached 96.6%, demonstrating excellent generalization. Training and validation loss curves indicated stable learning behavior with minimal overfitting.

Evaluation Metrics and Results

The evaluation was conducted using confusion matrices,accuracy,precision and recall for training and validation sets.The confusion matrix highlighted minor misclassification between benign and mailgnant classes while the rest are classified consistently with high precision

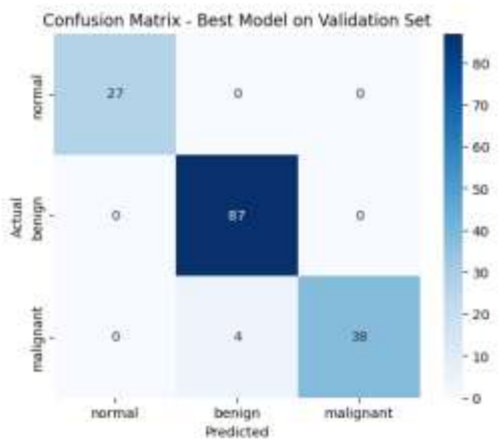

Fig 7.Validation confusion matrix for TumorNet

Stage 2.2: Shape Classification using DeepShapeNet

The shape classification model, DeepShapeNet, was a custom-designed Convolutional Neural Network that took binary tumor masks (64×64) as input and predicted one of four shape categories.

The network featured three convolutional blocks with dropout, batch normalization, and ReLU activations, followed by two fully connected layers for classification.

VOLUME: 09 ISSUE: 06 | JUNE - 2025 SJIF RATING: 8.586 ISSN: 2582-3930

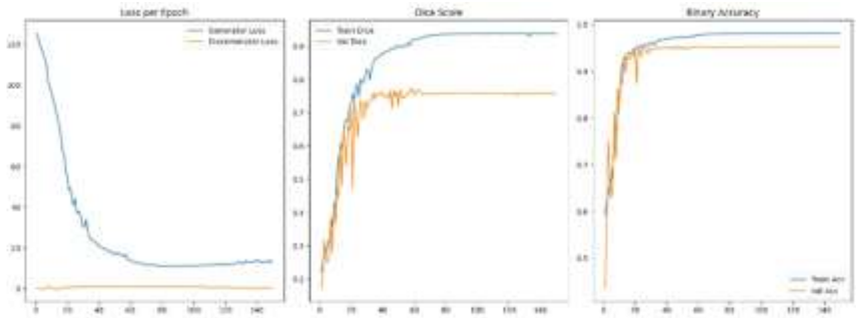

Fig 8. Accuracy Graphs

The model was trained using categorical cross-entropy loss and the Adam optimizer with a learning rate of 1e-3. Early stopping was used based on validation accuracy. The training achieved 98.4% accuracy, while the validation accuracy reached 94.6%, demonstrating excellent generalization. Training and validation loss curves indicated stable learning behavior with minimal overfitting.

Evaluation Metrics and Results

The evaluation was conducted using confusion matrices, accuracy, precision, recall, and F1-score for both training and validation datasets. The confusion matrix for the validation set highlighted minor misclassifications between irregular and lobular shapes due to their similar contours, while oval and round shapes were consistently classified with high precision.

Accuracy vs. epoch and loss vs. epoch graphs were plotted to observe training behavior, confirming smooth convergence. The segmentation masks used in classification preserved essential shape features, validating the strength of the pipeline.

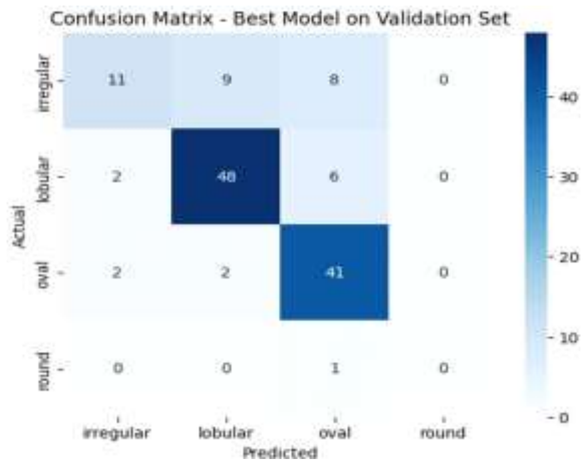

Fig 9 . Validation Set Confusion Matrix for DeepShapeNet

Pipeline Inference on New Samples

A complete inference pipeline was evaluated by feeding a new ultrasound image to the Stage I Attention U-Net, generating a tumor mask, which was then classified by the CNN. The model successfully predicted the correct tumor type and tumor shape with high confidence, showcasing real-time applicability. Visual outputs displayed the original image, predicted mask, and final shape label, making the pipeline interpretable and practical for clinical settings.

CONCLUSION

This study presented a two-stage deep learning pipeline for automatic tumor shape classification from breast ultrasound images. In Stage I, an enhanced **Attention U-Net** was employed to accurately segment tumor regions from noisy grayscale input, using attention mechanisms to focus on relevant spatial features.In Stage 2.1 the segmented masks are classifed using a **CNN(TumorNet)** into three tumor type categories:*normal,benign,malignant*.In Stage 2.2, the segmented masks were classified using a custom- designed **CNN (DeepShapeNet)** into four clinically significant shape categories: *irregular*,*lobular*, *oval*, and *round*.

The proposed system achieved high performance across all stages, with segmentation outputs closely aligning with ground truth masks and the shape classifier attaining **94.6% validation accuracy**. The modularity of the pipeline makes it adaptable to other medical image segmentation-classification tasks. Moreover, the interpretability of visual outputs and class labels enhances its practical utility in aiding radiologists during diagnosis and treatment planning.

In future work, we plan to extend this pipeline to multi- modal medical imaging (e.g., mammography and MRI) and explore transformer-based architectures for improved segmentation. Further, integrating clinical metadata (e.g., age, BI-RADS score) could enhance diagnostic accuracy. Real-world deployment and validation with radiologists will also be essential for clinical translation.

REFERENCES

- R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2017,” *CA: A Cancer Journal for Clinicians*, vol. 67, no. 1, pp. 7–30, 2017.

- B. Lauby-Secretan *et al.*, “Breast-cancer screening—viewpoint of the IARC working group,” *New England Journal of Medicine*, vol. 372, no. 24, pp. 2353–2358, 2015.

- R. M. Rangayyan, S. Banik, and J. L. Desautels, “Computer-aided detection of architectural distortion in prior mammograms of interval cancer,” *J. Digit. Imaging*, vol. 23, no. 5, pp. 611–631, 2010.

- J. Tang *et al.*, “Computer-aided detection and diagnosis of breast cancer with mammography: recent advances,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 2, pp. 236–251, 2009.

- Z. Jiao, X. Gao, Y. Wang, and J. Li, “A parasitic metric learning net for breast mass classification based on mammography,” *Pattern Recognit.*, vol. 75, pp. 292–301, 2018.

- V. K. Singh *et al.*, “Conditional generative adversarial and convolutional networks for x-ray breast mass segmentation and shape classification,” in *Proc. MICCAI*, Granada, Spain, 2018, pp. 833–840.

- J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Netw.*, vol. 61, pp. 85–117, 2015.

- J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 3431–3440, 2015.

- O. Ronneberger, P. Fischer, and T. Brox, “U- net: Convolutional networks for biomedical image segmentation,” in *Proc. MICCAI*, pp. 234– 241, 2015.

- V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A deep convolutional encoder- decoder architecture for image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, 2017.

- G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.

- M. A. Al-antari *et al.*, “A fully integrated computer-aided diagnosis system for digital x-ray mammograms via deep learning detection, segmentation, and classification,” *Int. J. Med. Inform.*, vol. 117, pp. 44–54, 2018.
- P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Proc. CVPR*, pp. 5967– 5976, 2017.
- I. Goodfellow *et al.*, “Generative adversarial nets,” in *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 2672–2680, 2014.
- M. M. K. Sarker *et al.*, “SLSDeep: Skin lesion segmentation based on dilated residual and pyramid pooling networks,” in *Proc. MICCAI*, Granada, Spain, 2018, pp. 21–29.