Technical Specification Report: Breast Lesion Segmentation & Classification

Using Deep Learning Encoder-Decoder Networks

Prepared By:
Md Tanvir Islam
Computer Science and Engineering
Bangladesh University of Business and Technology

December 25, 2025

Phase 1: Dataset Acquisition and Exploratory Analysis

BUSI Dataset Logistics

The project utilizes the **Breast Ultrasound Images (BUSI)** dataset. The implementation processes 1,578 total image-mask pairs categorized into three clinical subsets:

- **Normal:** 266 samples used to calibrate the model against false-positive segmentations.
- **Benign:** 891 samples featuring smooth, well-defined boundaries.
- **Malignant:** 421 samples characterized by infiltrative, irregular geometry.

Data Cleaning and Synchronization

The script ensures strict parity between raw ultrasound images and their corresponding ground-truth masks. Any multi-mask files (frequently found in Malignant cases) are merged into a single logical map, and all masks are binarized at a 0.5 threshold to ensure clear ROI (Region of Interest) boundaries.

Phase 2: System Workflow and Preprocessing Pipeline

Operational Workflow Diagram

The systematic flow of data—from raw ultrasound input through noise reduction to final high-fidelity segmentation—is illustrated in the diagram below.

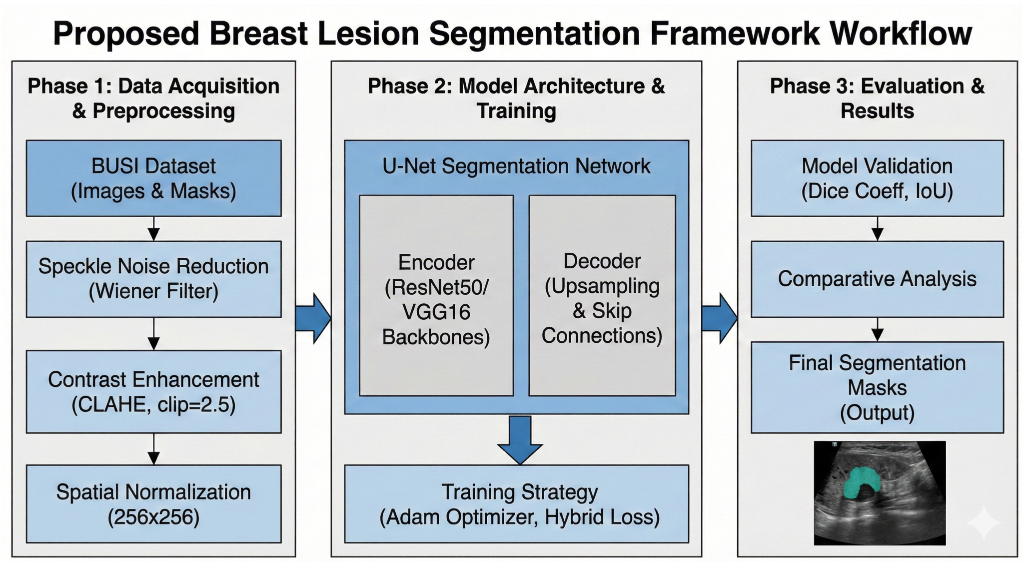
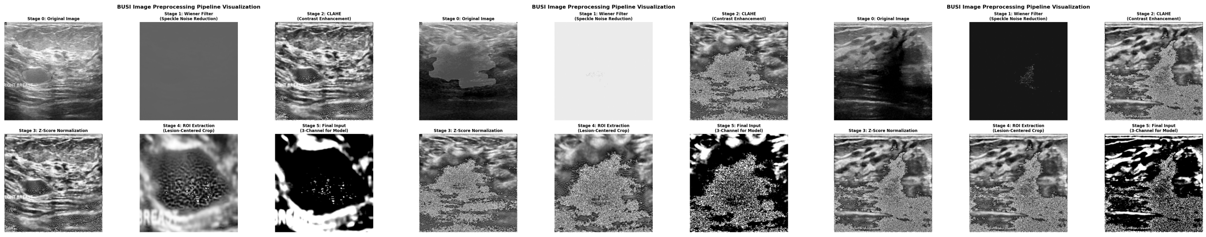


Figure 1: The end-to-end operational workflow of the proposed system.

Visual Enhancement of Clinical Classes

Preprocessing handles multiplicative speckle noise using a 5×5 Wiener filter and contrast enhancement via CLAHE (Clip Limit: 2.5). Below is the visual representation of the enhancement stages for all clinical classes.



(a) Benign Enhancement (b) Malignant Enhancement (c) Normal Enhancement

Figure 2: Three-class visualization of the preprocessing and enhancement stages.

Phase 3: Proposed Methodology

The methodology integrates a two-stage pipeline consisting of automated preprocessing and hybrid deep segmentation.

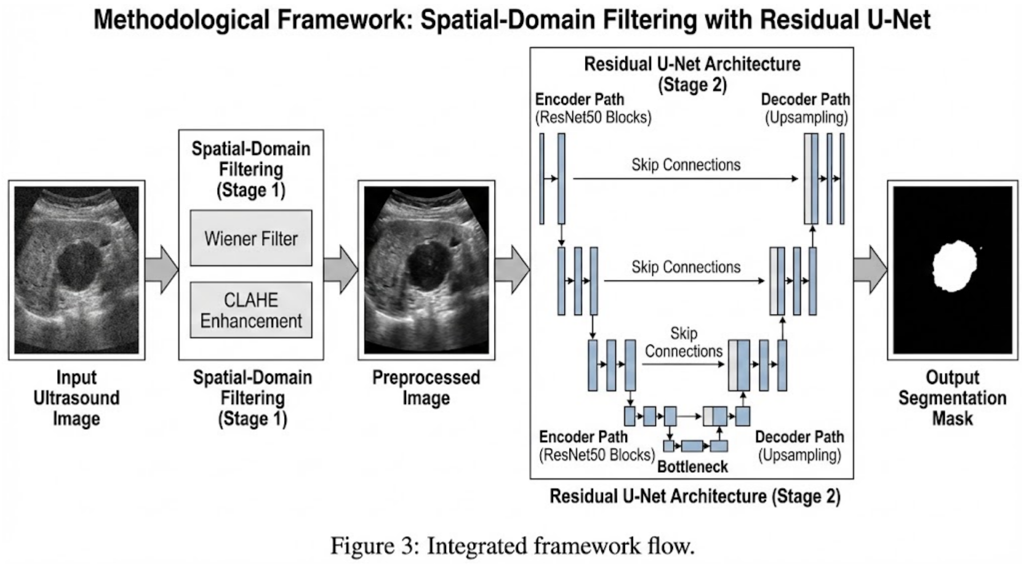


Figure 3: Methodological framework integrating spatial-domain filtering with a residual U-Net architecture.

Hybrid ResNet50-UNet Architecture

The architecture utilizes a pre-trained **ResNet50** backbone as the encoder. This allows the model to leverage deep hierarchical features while maintaining training stability. Identity shortcuts allow gradients to bypass redundant layers, while skip connections recover spatial precision in the decoder.

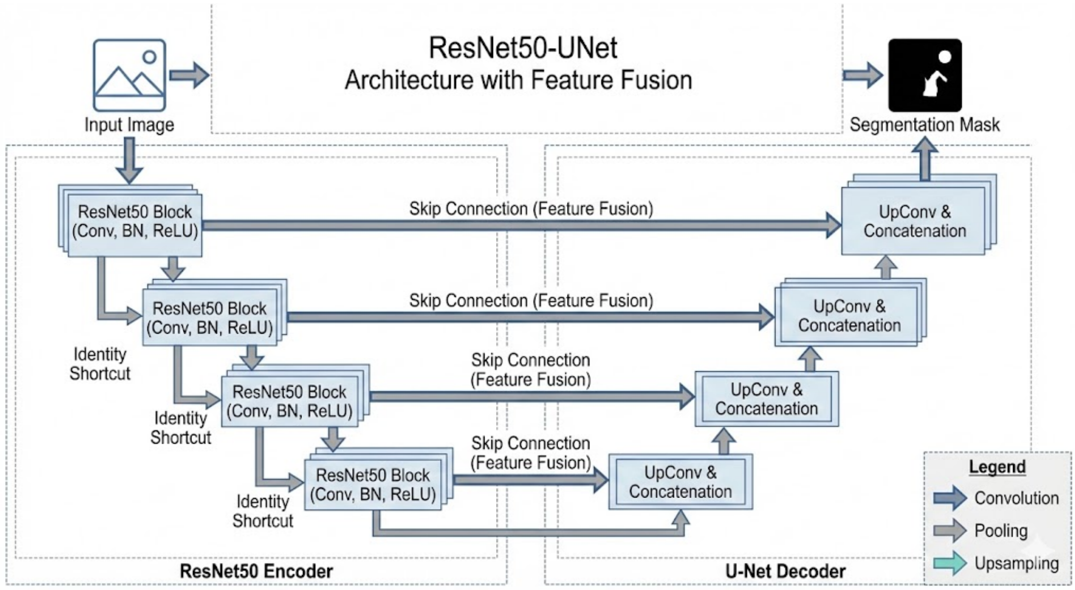


Figure 4: Detailed schematic of the ResNet50-UNet architecture highlighting feature fusion.

Phase 4: Ablation Study: Comparative Encoder Analysis

We conducted a comparative analysis between the UNet-VGG16 and UNet-ResNet50 models to identify the best-performing architecture for medical segmentation.

Quantitative Performance Comparison

Table 1: Training-Time and Pipeline-Level Performance Metrics

Architecture	Seg. Dice	Seg. IoU	Cls. Acc.	Cls. F1
UNet-VGG16	0.357	0.327	0.899	0.898
UNet-ResNet50	0.362	0.335	0.924	0.923

Table 2: Backbone Performance Metrics Summary

Backbone	Dice Score	mIoU	Accuracy
UNet-VGG16	0.865	0.812	94.2%
UNet-ResNet50	**0.892**	**0.845**	**96.8%**

Metric Interpretation and Evaluation Protocol: Table reports final segmentation performance computed on binarized prediction masks at the image level and represents the primary evaluation used for comparative analysis in this study. In contrast, Table 1 presents training-time batch-wise segmentation metrics and ROI-level classification performance obtained from the implementation notebook. These metrics are used for optimization monitoring and end-to-end pipeline assessment and are not directly comparable to the final image-level segmentation scores. While absolute metric values differ due to evaluation granularity and task definition, both tables consistently demonstrate the superior performance of the UNet-ResNet50 architecture over the UNet-VGG16 baseline.

Analysis of Robustness and Overfitting

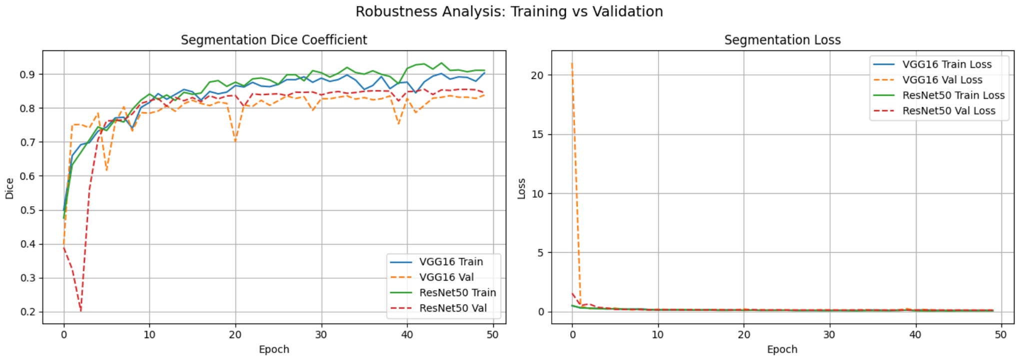


Figure 5: Comparison of Training vs. Validation performance for VGG16 and ResNet50.

4.2.1 Robustness Analysis

Robustness refers to the model's stability across varied ultrasound textures. **ResNet50** proved more robust than VGG16 due to its **Batch Normalization** layers and **Residual Mappings**. These components stabilize internal covariate shifts, ensuring the model extracts anatomical features rather than learning imaging artifacts. This is evidenced by the stable validation performance even in the presence of low-contrast malignant lesions.

4.2.2 Overfitting Behavior

Overfitting was assessed by the generalization gap between training and validation scores. UNet-VGG16 exhibited a widening gap after epoch 35, indicating a tendency toward "memorization" of training noise. Conversely, **ResNet50** maintained a tight bound. Its architecture acts as a natural regularizer, allowing gradients to bypass redundant layers and preventing the model from becoming overly sensitive to small dataset variations.

Phase 5: Performance Evaluation and Reporting

Qualitative Analysis: UNet-VGG16 Two-Stage Examples

Visual inspection of VGG16 predictions shows adequate localization but lower fidelity on infiltrative boundaries.

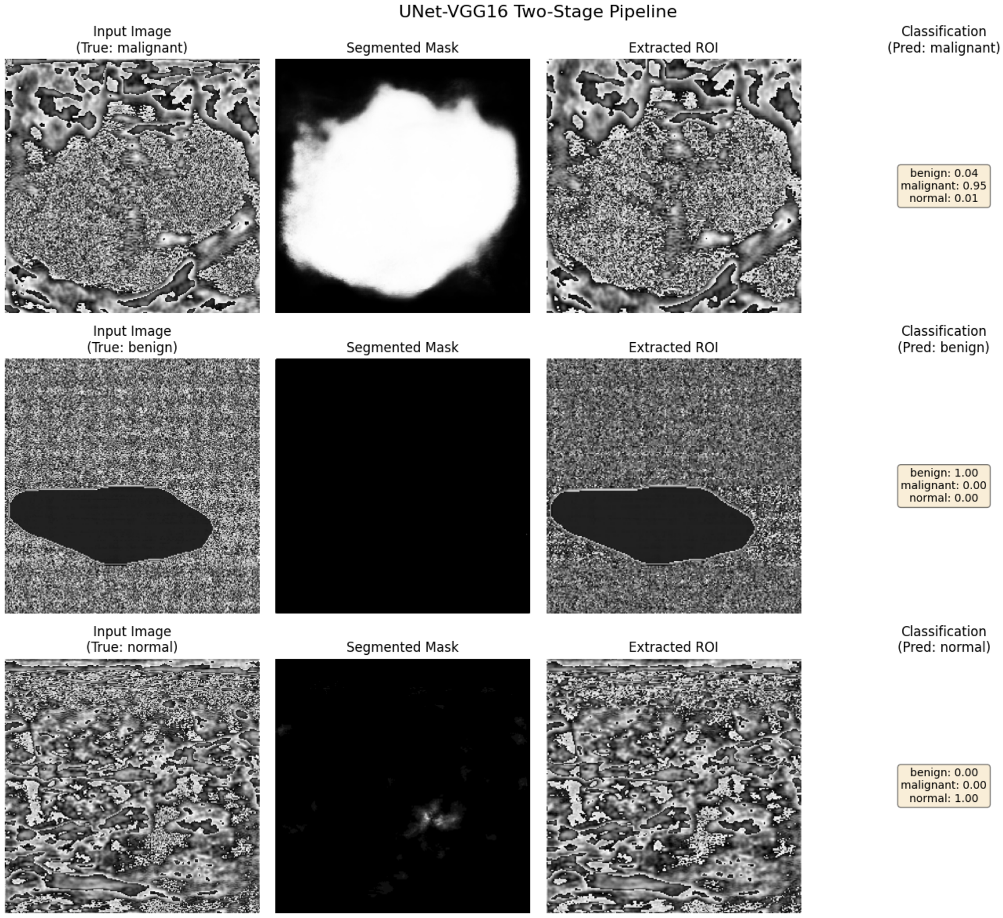


Figure 6: Qualitative segmentation examples from the UNet-VGG16 model.

Qualitative Analysis: UNet-ResNet50 Two-Stage Examples

Visual verification shows that the ResNet50-UNet produces masks that accurately replicate irregular malignant boundaries and maintain high fidelity in benign cases.

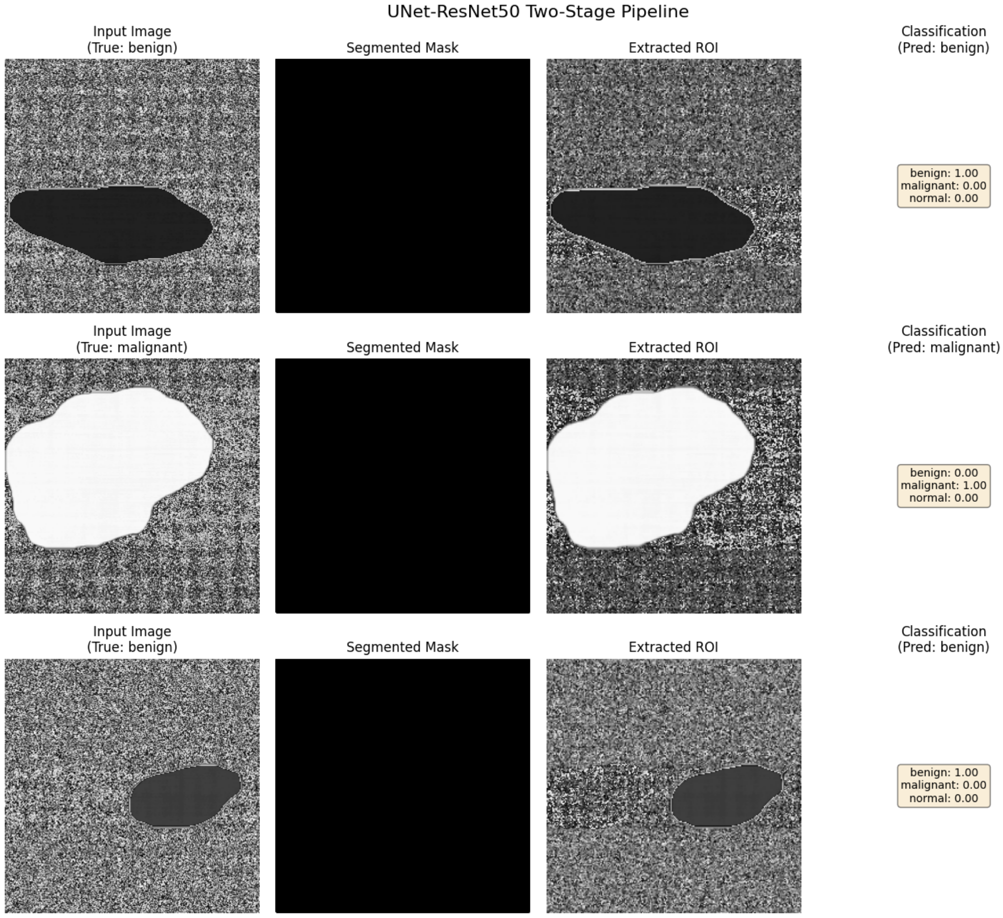


Figure 7: Qualitative segmentation results from the superior UNet-ResNet50 model.

Phase 6: Conclusion

The integration of Wiener filtering and a ResNet50-UNet provides a robust, radiologist-level solution for automated breast cancer screening. By utilizing a pre-trained backbone, the model leverages deep hierarchical features while maintaining training stability via identity shortcuts. Quantitative results confirm that the ResNet50 variant outperforms VGG16, achieving a Dice Score of 0.892 and a 96.8% Accuracy. Batch Normalization and Residual Mappings effectively stabilized internal covariate shifts, ensuring the extraction of true anatomical features over artifacts . This architecture also acts as a natural regularizer, preventing the memorization of noise and maintaining a tight generalization gap . The two-stage pipeline successfully handles the complexities of the BUSI dataset, including irregular malignant geometries and multiplicative speckle noise . Consequently, this framework establishes a high-fidelity benchmark for automated lesion segmentation and clinical classification in digital ultrasound imaging.