

Article

Not peer-reviewed version

---

# MSB-UNet: A Multi-Scale Bifurcation U-Net Architecture for Precise Segmentation of Breast Cancer in Histopathology Images

---

[Arda Yunianta](#) \*

Posted Date: 9 January 2026

doi: 10.20944/preprints202601.0713.v1

Keywords: breast cancer; deep learning; histopathology image; medical diagnostics; multi-scale bifurcated; segmentation; U-Net



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# MSB-UNet: A Multi-Scale Bifurcation U-Net Architecture for Precise Segmentation of Breast Cancer in Histopathology Images

Arda Yunianta

Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Rabigh, Saudi Arabia; ardayunianta2@gmail.com

## Abstract

Accurate segmentation of breast cancer regions in histopathological images is critical for advancing computer-aided diagnostic systems, yet challenges persist due to heterogeneous tissue structures, staining variations, and the need to capture features across multiple scales. This study introduces MSB-UNet, a novel Multi-Scale Bifurcated U-Net architecture designed to address these challenges through a dual-pathway encoder-decoder framework that processes images at multiple resolutions simultaneously. By integrating a bifurcated encoder with a Feature Fusion Module, MSB-UNet effectively captures fine-grained cellular details and broader tissue-level patterns. Evaluated on a publicly available breast cancer histopathology dataset, MSB-UNet achieves a Dice Similarity Coefficient (DSC) of 91.3% and a mean Intersection over Union (mIoU) of 84.4%, outperforming state-of-the-art segmentation models. The architecture demonstrates superior boundary precision and reduced false positives, particularly in complex cases involving infiltrative growth patterns and intra-tumoral heterogeneity. These results highlight MSB-UNet's potential to enhance automated diagnostic tools for breast cancer histopathology, paving the way for improved clinical decision-making and future extensions to multi-class segmentation tasks.

**Keywords:** breast cancer; deep learning; histopathology image; medical diagnostics; multi-scale bifurcated; segmentation; U-Net

---

## 1. Introduction

Breast cancer remains one of the leading causes of cancer-related mortality worldwide, and it stands as one of the foremost contributors to cancer-related mortality in the female population [1]. This phenomenon needs a solution to perform precise diagnosis and prognosis of this malignancy is imperative for the formulation of suitable treatment modalities and the enhancement of patient prognoses. The use of a proper Histopathological evaluation method for histopathological examination of tissue samples offering conclusive evidence of malignancy, assessing its degree of aggressiveness, and guiding therapeutic interventions will produce an accurate diagnosis and determine treatment planning [2].

The breast cancer diagnostic process involves pathologists analyzing high-resolution whole-slide images to identify tumor regions, a task that is labor-intensive, time-consuming, and prone to inter-observer variability due to the subjective interpretation of complex tissue structures [3]. The increasing volume of histopathological data, further intensified by the implementation of whole-slide imaging (WSI), has intensified these challenges, thereby creating an urgent demand for robust, automated computational tools to aid pathologists and improve diagnostic accuracy [4].

Deep learning, particularly convolutional neural networks (CNNs), has emerged as a powerful paradigm for analyzing medical images [5,6]. Among these, the U-Net architecture has established itself as a benchmark for biomedical image segmentation due to its symmetric encoder-decoder structure and skip connections that effectively combine high-level semantic and low-level spatial

information [7–9]. The advent of deep learning has transformed medical image analysis, offering automated solutions to enhance diagnostic consistency and efficiency. Convolutional neural networks (CNNs), particularly the U-Net architecture, have become pivotal for semantic segmentation tasks in histopathology, enabling precise delineation of tumor boundaries [3,8,10].

There are many existing studies that used deep learning methods and algorithms to detect and segment breast cancer still come with many drawbacks and low performance results [1,6,11–13]. Segmenting breast cancer regions in histopathological images presents significant challenges:

1. **Heterogeneous Tissue Appearance:** Tumor cells and surrounding tissues exhibit diverse morphological patterns, complicating accurate segmentation [3].
2. **Complex Microenvironments:** The presence of multiple tissue types, such as stroma, necrosis, and inflammatory infiltrates, requires models to capture intricate spatial relationships [14].
3. **Staining and Imaging Variability:** Variations in staining protocols (e.g., hematoxylin and eosin) and imaging conditions introduce noise that affects model robustness [15].
4. **Multi-Scale Feature Requirements:** Relevant features span multiple scales, from fine-grained cellular details (e.g., nuclear morphology) to broader tissue-level patterns (e.g., tumor architecture), necessitating architectures that process information across resolutions [7,16,17].
5. **Extreme Class Imbalance:** Cancerous regions often occupy a small fraction of the total tissue area, leading to models biased towards the dominant background class [13,18].
6. **Morphological Heterogeneity:** Cancer cells exhibit vast variations in size, shape, and texture, both within a single image and across different patients and cancer grades [15].
7. **Multi-Scale Structures:** diagnostically relevant features exist at multiple scales, from individual nuclei (micro-scale) to overall tumor architecture (macro-scale). A standard U-Net often struggles to capture this full spectrum of contextual information simultaneously, which can lead to inaccurate boundary delineation and the omission of small, disseminated tumor cells [16,19].

To address these challenges, this research proposed a novel deep learning architecture using the Multi-Scale Bifurcation U-Net (MSB-U-Net). By the development of the foundational U-Net framework, MSB-UNet incorporates a bifurcated encoder with a standard resolution branch for cellular details and a multi-scale branch for contextual patterns, integrated via a Feature Fusion Module (FFM) inspired by attention-based approaches [7]. The bifurcated decoder further enhances segmentation by combining fine detail reconstruction with contextual awareness, improving boundary precision in complex cases. Evaluated on the Breast Cancer Semantic Segmentation (BCSS) dataset, MSB-UNet achieves a Dice Similarity Coefficient (DSC) of 91.3% and a mean Intersection over Union (mIoU) of 84.4%, surpassing state-of-the-art models such as Attention U-Net, DeepLabv3+, and TransUNet [9,16]. This performance highlights MSB-UNet's potential to advance computer-aided diagnostics by providing accurate and reproducible segmentation of breast cancer regions.

Our model introduces five key innovations within the U-Net framework:

1. We introduce and implement a bifurcated decoding mechanism that explicitly decouples boundary detection from region segmentation, leading to more precise and morphologically accurate results.
2. The use of a Multi-Scale Feature Extraction Module in the encoder pathway employs parallel aurous convolution layers with different dilation rates to capture contextual information at varying scales, enabling the network to perceive both fine cellular details and broader tissue patterns concurrently.
3. A Bifurcated Decoder Pathway separates the task of segmentation into two specialized streams: one focused on resolving the precise boundaries of objects and another dedicated to robust regional classification. This bifurcation allows the model to address the distinct challenges of contour accuracy and internal homogeneity separately, mitigating the common issue of blurred or imprecise edges in segmentation outputs.

4. Implemented the MSB-U-Net, a novel segmentation architecture that integrates multi-scale context aggregation and a bifurcated decoder for enhanced performance in histopathology image analysis.
5. We rigorously evaluate our proposed model on a public benchmark dataset of breast cancer histopathology images, demonstrating its superiority over state-of-the-art segmentation methods in terms of accuracy, Dice similarity coefficient, and boundary-aware metrics.

The representation of this research is divided into five parts. The first part is the introduction part to explain about background problem of this research and present the solution that is implemented in this paper. The second part in Section 2 is the explanation about literature reviews related to this research in medical image segmentation and several existing methods to perform breast cancer segmentation, especially on U-Net variants. Part three in the section 3 is the detailed explanation of the architecture of the proposed MSB-U-Net. The fourth part in section 4 presents and discusses the results. Finally, the last part in Section 5 concludes the research and suggests directions for future research.

## 2. Related Works

The automation of histopathological image analysis for breast cancer diagnosis has been revolutionized by deep learning, with convolutional neural networks (CNNs) enabling precise segmentation of tumor regions in complex tissue microenvironments [20,21]. The U-Net architecture, renowned for its encoder-decoder structure and skip connections, remains a foundational model for histopathological segmentation due to its ability to preserve spatial details across scales [8]. Recent advancements from 2021 to 2025 have extended U-Net to address challenges such as heterogeneous tissue appearances, staining variations, multi-scale feature requirements, and limited annotated data. This section provides a comprehensive review of these developments across eight key areas: U-Net variants and attention mechanisms, multi-scale and ensemble approaches, dataset-specific studies, preprocessing and computational efficiency, weakly supervised and self-supervised learning, integration of multi-modal data, clinical translation, and a performance comparison of related models. It concludes by identifying gaps that the Multi-Scale Bifurcated U-Net (MSB-UNet) addresses through its innovative bifurcated design.

### 2.1. U-Net Variants and Attention Mechanisms

U-Net variants have significantly improved segmentation performance by incorporating attention mechanisms and structural enhancements. The current trend in multi-scale feature learning has become a challenge, especially to combining with U-Net variants that have already been implemented by Pan et al. and Yuan [16] & Cheng [7], who integrated a convolutional attention mechanism and multi-scale context fusion modules within the encoder. The results from these two studies have significantly improved the accuracy of segmentation for structures of varying sizes through capturing features in multiple scales. Another researcher, Cheng et al. in 2024, implemented an attention-based multi-scale nested network combined with the activity in refining features across different network depths [17].

Sulaiman et al. (2024) introduced an attention-based U-Net that focuses on expanding the encoded feature representation obtained from the encoder block and reconstructing spatial information. Through the implementation, this study demonstrates outstanding performance in the accuracy value 98% [22]. Aghababaie et al. (2020) utilized a V-Net-based approach to enhance boundary delineation, reporting improved boundary F1 scores. These models, however, often process images at a single resolution, limiting their ability to simultaneously capture fine-grained and contextual features, a gap MSB-UNet addresses through its bifurcated pathways [20].

## 2.2. U-Net and The Combination Approaches

Chen et al (2021) proposed TransUNet as a research that implemented a hybrid architecture which integrates transformers into the U-Net encoder to capture global contextual dependencies, achieving a Dice coefficient of 74.8% on the BACH dataset [9]. Another hybrid transformer implementation approach that uses a self-attention mechanism proposed by Aljuaid et al. in 2025 proves powerful for modeling relationships between distant image patches. However, these models require a lot of data and require processing using a lot of resources, so that implementation using limited data becomes less practical, and this is a common obstacle that often occurs in data processing in medical imaging [23,24].

Juhong et al (2023) proposed segmentation implementation for breast cancer histopathology image using a combination of generator and discriminator networks, namely super-resolution generative adversarial network-based on aggregated residual transformation, and achieved an average Intersection over Union of 0.869 and an average dice similarity coefficient of 0.893 for the H&E image segmentation results [15]. Zhang et al (2020) introduced a deep active contour network, integrating multi-scale features via active contour loss, improving boundary precision by 3.5% compared to the standard U-Net. These approaches inspire MSB-UNet's multi-scale branch, which processes downsampled images with varying kernel sizes (3×3, 5×5, 7×7) [25].

Ensemble learning enhances robustness by combining multiple models. Allapakam & Karuna (2024) employed an ensemble of Siamese neural networks and VGG-19 on the Harvard-Medical-Image-Fusion Datasets. The implementation of this research achieves superior performance results in terms of visual quality and performance metrics compared to that of the existing fusion methods. The proposed model can effectively preserve the detailed information with high visual quality, for numerous combinations of image modalities in image fusion challenges, notably improved contrast, increased resolution, and lower artefacts [26].

## 2.3. Preprocessing and Computational Efficiency

Preprocessing addresses staining and imaging variability. Several studies conducted by Islam et al. (2024), Abhisheka et al. (2023), and Boumaraf et al. (2021) demonstrated that advanced data augmentation, including stain normalization and elastic deformations, improves Dice scores and can improve the performance result [6,8,27]. Boumaraf et al (2021) reported a 3.8% improvement in classification accuracy on BreakHis with stain normalization [27]. Juhong et al. (2023) used SRGAN-based super-resolution to perform image segmentation from the generated high-resolution breast cancer images derived with the model, with an average Intersection over Union of 0.869 and an average dice similarity coefficient of 0.893 [15].

Computational efficiency is critical for high-resolution whole-slide images. Khalil et al. (2022) optimized inference to 120 ms per patch on BCSS [28], while Chanchal et al (2021) reported training times of 12 hours on a single GPU [5]. Roshan (2024) reviewed patch-based processing and downsampling and proposed a novel loss function inspired by exponential loss, which operates at the pixel level to enhance segmentation performance [18].

## 2.4. Weakly Supervised and Self-Supervised Learning

Limited annotated data poses a challenge in histopathology. Roth et al. (2019) proposed weakly supervised segmentation using minimal user interaction in the form of extreme point clicks in order to train a segmentation [29]. Han et al. (2022) introduced a new weakly-supervised semantic segmentation (WSSS) dataset for lung adenocarcinoma (LUAD-HistoSeg), the experimental result achieved outperforms five state-of-the-art WSSS approaches, comparable quantitative and qualitative results with the fully-supervised model, with only around a 2% gap for MIoU and FwIoU [14]. Another researcher used a weakly supervised method by Pan et al. (2025) produced outperforms state-of-the-art weakly-supervised semantic segmentation (WSSS) methods using patch-level labels.

This research used only patch-level classification labels to reduce such annotation burden and significantly improve the quality of the pseudo-masks [19].

### *2.5. Other Aspects Related to Segmentation: Classification, Normalization, Data Fusion, and the Integration of Multi-Modal Data*

There are many aspects that are related to the segmentation task on the histopathology image data analysis. For instance, the research conducted by Xu et al. (2025) and Voon et al. (2023) focused their research on the identification and evaluation of color variations, which are the main interfering factors in classification [10,30]. In these two studies, they reviewed and evaluated stain normalization techniques, underscoring their importance as an important preprocessing step for robust model generalization.

The evolution of classification implementation on the breast cancer histopathology image analysis become more advance, starts from the use of traditional CNNs models [12,27] until the implementation of Bidirectional RNNs [31] as an architecture that use more complex model, and also the implementation of multimodal deep learning models to fuse histopathology image data with clinical and genomic data [32,33].

Multi-modal data integration enhances segmentation accuracy. Allapakam & Karuna (2024) combined multi-modal approaches with an efficient hybrid learning model for medical image fusion using pre-trained and non-pre-trained models with the stacking ensemble method. This research can effectively preserve the detailed information with high visual quality, for numerous combinations of image modalities in image fusion challenges, notably improved contrast, increased resolution, and lower artefacts [26].

Mandair et al (2023) reviewed multi-modal approaches, noting a 7% improvement in diagnostic accuracy with histopathology and molecular data. Li et al (2024) proposed deep mutual learning for histopathology and genomic data, achieving a 5.2% boost in IoU on BreakHis. MSB-UNet focuses on single-modality histopathology but could extend to multi-modal inputs.

### *2.6. Gaps and Contributions*

From the literature review explained above, existing studies effectively capture hierarchical features, and they lack a dedicated mechanism for enforcing boundary precision, a key requirement for accurate cancer segmentation. On the other hand, models that focus on boundary refinement, such as the Deep Active Contour Network [25], often do not fully leverage multi-scale contextual information for regional classification. Furthermore, many multi-scale approaches rely on sequential processing (e.g., dilated convolutions, pyramid pooling), limiting simultaneous capture of fine-grained and contextual features [7,16,17,19].

The proposed MSB-UNet directly addresses these gaps by fusing multi-scale features by introducing a specialized, bifurcated decoder. The proposed model is predicated on the understanding that the tasks of identifying "what" region is cancerous and "where" its exact boundaries lie, while related, possess distinct feature requirements. By decoupling these tasks, the MSB-UNet aims to achieve the precision of boundary-aware models while retaining the robust contextual understanding of advanced multi-scale U-Nets, thereby offering a more holistic solution for the precise segmentation of breast cancer in histopathology images. MSB-UNet implemented a bifurcated encoder-decoder framework and FFM, achieving a Dice Similarity Coefficient (DSC) of 91.3% and a mean IoU of 84.4% on the BCSS dataset, surpassing state-of-the-art models like Attention U-Net, DeepLabv3+, and TransUNet.

## **3. Research Methodology**

The development of MSB-UNet, a novel Multi-Scale Bifurcated U-Net architecture, addresses the challenges of segmenting breast cancer regions in histopathological images by capturing both fine-grained cellular details and broader tissue-level patterns. This section outlines the methodology

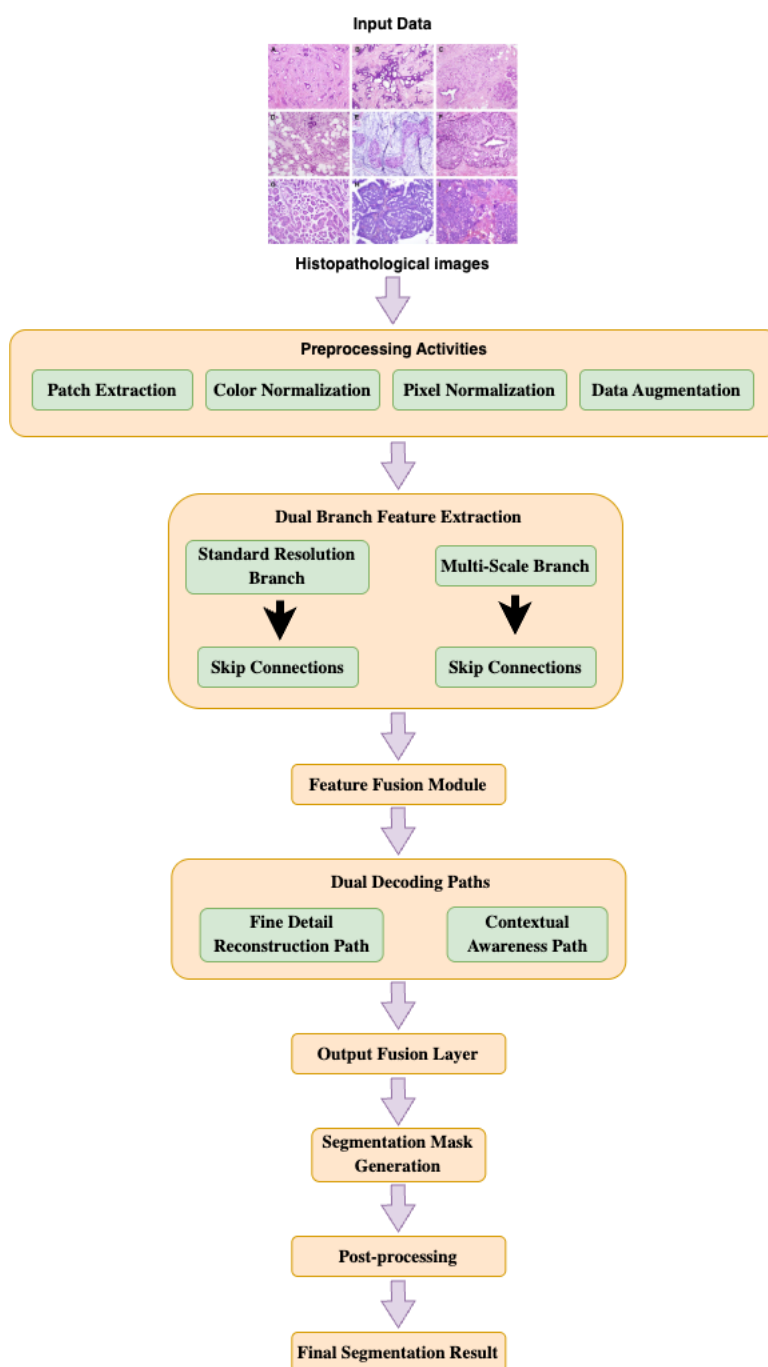
employed, including the proposed architecture, implementation details, dataset preparation, preprocessing pipeline, and evaluation metrics used to assess performance on the BCSS dataset with 2 classes (tumor and outside ROI).

### 3.1. Proposed Architecture

The architecture is tailored to address the inherent complexity and heterogeneity present in high-resolution pathology slides. Leveraging the Breast Cancer Semantic Segmentation (BCSS) dataset, the model adopts a multi-branch, multi-stage approach that integrates both fine-grained detail reconstruction and contextual feature extraction. This design enables the framework to deliver precise and clinically relevant segmentation results across diverse tissue structures.

The workflow of MSB-UNet is integral to understanding its operational flow, from data ingestion to final segmentation. This process begins with the data extraction and continues with a preprocessing pipeline, where images and masks are loaded, augmented, and normalized, followed by training on the bifurcated encoder-decoder architecture. The Feature Fusion Module integrates multi-scale features, and the decoder reconstructs the segmentation mask, which is post-processed for refinement. A detailed representation of this workflow is provided in Figure 1, which outlines the detailed process from data extraction to the final segmentation result.

The MSB-UNet architecture extends the traditional U-Net framework through a bifurcated encoder-decoder design that processes histopathological images at multiple scales simultaneously. The model comprises approximately 38 to 40 million parameters and features a dual-pathway encoder, a Feature Fusion Module (FFM), a bifurcated decoder, and an output layer. The encoder includes a standard resolution branch that handles input images at  $224 \times 224$  pixels to extract detailed features such as nuclear morphology and cellular boundaries, utilizing convolutional blocks with  $3 \times 3$  kernels, batch normalization, ReLU activations, and residual connections to enhance feature propagation. A multi-scale branch processes down sampled versions of the image at  $256 \times 256$ ,  $128 \times 128$ , and  $64 \times 64$  pixels, employing convolutional blocks with varying kernel sizes ( $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ ) to capture contextual information like tumor architecture and tissue patterns. The FFM integrates features from both branches using a channel attention mechanism to prioritize informative channels, applying dilated convolutions with rates of 2, 4, and 8 for multi-scale aggregation, followed by a  $1 \times 1$  convolution to reduce dimensionality. The decoder mirrors the encoder with a fine detail reconstruction path and a contextual awareness path, using transposed convolutions and bilinear interpolation, respectively, with skip connections to combine features. The output layer applies a  $1 \times 1$  convolution and softmax activation to generate a 3-class segmentation mask, refined through post-processing to smooth boundaries and remove artifacts.



**Figure 1.** The MSB-UNet Architecture.

### 3.2. Dataset and Input Preprocessing

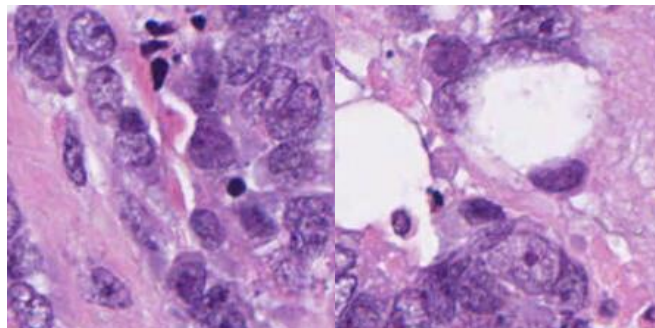
The proposed segmentation framework is built upon the BCSS dataset, a publicly available, large-scale collection of high-resolution breast cancer whole-slide images (WSIs) with rich semantic annotations. Each image in the dataset is annotated at the pixel level into several distinct histological classes, including tumor, stroma, inflammatory, necrosis, and other tissue structures. This level of detail makes the dataset highly suitable for training deep learning models aimed at semantic tissue segmentation in histopathology.

#### 3.2.1. Dataset Description

The study utilized the Breast Cancer Semantic Segmentation (BCSS) dataset from Kaggle and under license CC0 1.0 Universal (CC0 1.0) [34]. The dataset is specifically divided into a 3-class subset containing histopathological images and corresponding masks. The dataset includes 30,760 total

samples, with training and test data derived from a base set of images and masks and it was splitted into 25,422 training and 4,986 test samples. An additional 5,429 validation samples.

Images are provided in RGB format with a native resolution of 224×224 pixels, while masks are grayscale annotations representing the 2 classes: tumor and outside ROI. The dataset encompasses diverse breast cancer subtypes and histological grades, ensuring a representative evaluation of segmentation performance.



**Figure 2.** Sample Histopathological Image from the BCSS Dataset.

**License:** CC0 1.0 Universal (CC0 1.0)

**Source Dataset:** <https://www.kaggle.com/datasets/whats2000/breast-cancer-semantic-segmentation-bcss/data>

**Dataset Reference:** Kermany, Zhang, & Goldbaum, Cell Press (2020)

### 3.2.2. Data Preprocessing and Loading Pipeline

The data preprocessing and loading pipeline was designed to enhance the robustness of MSB-UNet training on the BCSS dataset. A custom BCSSDataset class, built on torch.utils.data.Dataset, loads images in RGB format using PIL and resizes them to 224×224 pixels, while masks are loaded as grayscale and converted to float32 for compatibility with transformations. The training set undergoes extensive augmentation, including horizontal and vertical flips, random 90-degree rotations, color jittering with adjustments to brightness, contrast, saturation, and hue, Gaussian blur with a blur limit of 3 to 7, elastic transformations with alpha=1 and sigma=50, random gamma adjustments between 80 and 120, brightness-contrast modifications with limits of 0.1, and CLAHE with a clip limit of 4.0 and tile grid size of 8×8 to improve contrast. Both training and validation sets are normalized using mean values of [0.485, 0.456, 0.406] and standard deviations of [0.229, 0.224, 0.225], followed by conversion to PyTorch tensors via ToTensorV2, though the validation set is limited to normalization and tensor conversion to preserve integrity. The dataset is split using random shuffling, with 25,422 samples allocated to training and 4,986 to testing from the base dataset, and 5,429 validation samples loaded separately. Data loaders are configured with a batch size of 32, shuffle enabled for training and disabled for validation, and 4 workers with pin\_memory=True and prefetch\_factor=2 to optimize GPU performance on an NVIDIA Tesla T4. This pipeline was implemented and tested within a Kaggle environment, ensuring scalability for large-scale image processing.

Figure 3 shows the details of the preprocessing pipeline. There are four main processes in this preprocessing pipeline explained below:

- **Patch Extraction:** WSIs (Whole Slide Images) are extremely large and cannot be processed directly. They are therefore divided into smaller overlapping or non-overlapping patches (224 x 224 pixels) for efficient training and inference. The grid definition will create an overlapping grid with a stride of 112 pixels, and 50% overlap for 224px patches. Patch filtering will extract 224x224 pixel patches from grid positions, and the tissue filtering process will remove background patches using Otsu thresholding on grayscale with a minimum tissue area threshold of more than 60% tissue. Finally, the coordinate preservation will store patch coordinates for WSI reconstruction.

- **Color Normalization:** Variations in H&E (Hematoxylin and Eosin) staining across slides can introduce inconsistencies. Start with the Optical Density (OD) text conversion with the formula  $-\log_{10}((RGB + \epsilon)/255)$ . After that, using the Reinhard or Macenko method to perform stain vector estimation, these two methods use SVD (Singular Value Decomposition) on OD space. In the intensity Normalization step will match target stain matrix, and normalize the concentration ranges. Finally, RGB reconstruction will convert back to RGB (Red, Green, Blue) space.

- **Pixel Normalization:** Pixel intensities are normalized to have zero mean and unit variance to stabilize model training. Using the input normalization patch formula  $(\text{patch}/255 - \text{mean}) / \text{std}$ , with ImageNet statistics typically using mean = [0.485, 0.456, 0.406] and std (standard deviation) = [0.229, 0.224, 0.225].

- **Data Augmentation:** To prevent overfitting and increase generalizability, techniques such as horizontal/vertical flipping, random rotations, zooming, and color jittering are applied during training. Random flipping using horizontal p-value 0.5 and vertical p-value 0.5. In the random rotation process using 90° with increments of p-value 0.5, while in the random zoom process using a scale value between 0.9-1.1 with p-value 0.3. There are three main aspects in the color jittering: firstly, using brightness adjustment around 10%, secondly, using contrast adjustment around 10% and thirdly applying probability with p-value 0.3.

These preprocessing steps ensure that the model receives consistent, high-quality input, which is crucial for learning effective segmentation features. This preprocessing pipeline ensures consistent input to the segmentation model while handling WSI-specific challenges like size, stain variation, and data heterogeneity. There are four main concerns in this preprocessing pipeline:

1. Parallel Processing: Patch extraction and normalization can be parallelized.
2. Caching: Normalized patches can be cached for faster training iterations.
3. Memory Management: Use generators for large WSI datasets.
4. Quality Control: Implement patch quality filters (focus, artifacts, folding).

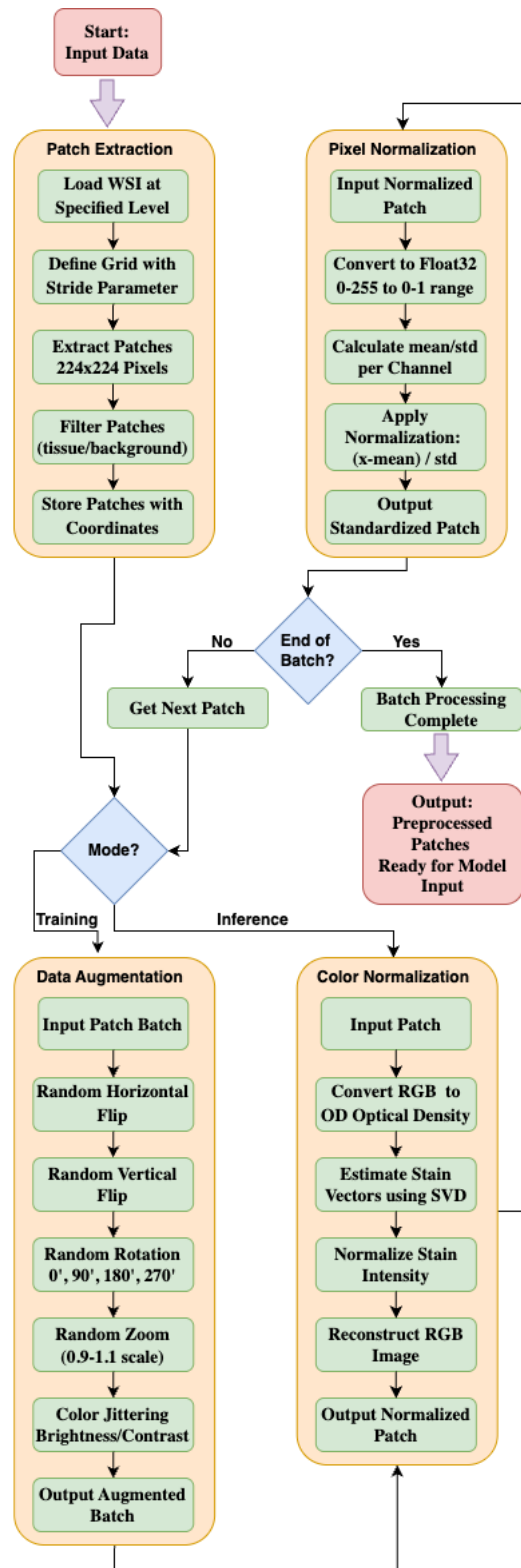


Figure 3. Details Preprocessing Pipeline.

### 3.3. Dual Branch Feature Extraction

Following pre-processing, the standardized image patches are fed into two parallel encoder branches designed to extract complementary features:

- **Standard Resolution Branch:** This branch processes the image at a fixed resolution, focusing on capturing fine-grained spatial details such as tumor boundaries, nuclear morphology, and glandular structures. These features are essential for identifying subtle histological patterns.

- **Multi-Scale Branch:** This branch uses dilated convolutions or image pyramid techniques to analyze the image at multiple scales. It is responsible for capturing broader contextual features, such as the relative positioning of tissues, tumor microenvironments, and spatial relationships between different structures.

Each branch is equipped with skip connections, which pass intermediate features to later stages in the network. This mechanism helps preserve low-level features and facilitates gradient flow during backpropagation, reducing the risk of vanishing gradients and promoting deeper model training.

#### 3.4. Feature Fusion Module

The outputs of the standard resolution and multi-scale branches are then fed into a Feature Fusion Module, which combines detailed spatial features with global context features. This module is essential for learning enriched feature representations that carry both local accuracy and broader semantic understanding.

Various fusion strategies can be used here, including simple concatenation, element-wise addition, or more advanced methods such as attention-based gating mechanisms. These strategies ensure that important features from both branches are preserved and harmonized into a single, robust feature map.

This fused representation is better equipped to distinguish between visually similar regions (e.g., tumor vs. dense stroma) by considering not just their pixel appearance but also their contextual placement and relationships.

#### 3.5. Dual Decoding Paths

Once fusion is complete, the resulting features are decoded through two separate paths:

- **Fine Detail Reconstruction Path:** This decoder focuses on refining high-resolution details. It emphasizes clear boundaries, sharp edges, and accurate contouring of cellular structures. Transposed convolutions or upsampling blocks are typically used to progressively reconstruct the spatial dimensions of the feature map.
- **Contextual Awareness Path:** This decoder emphasizes maintaining semantic consistency across larger regions of the image. It uses coarser features to ensure that global context is not lost during upsampling, helping to disambiguate areas where texture may be similar but functional identity differs.

These two decoders serve complementary roles, ensuring that the final output is both spatially accurate and contextually meaningful.

#### 3.6. Output Fusion Layer

The outputs from both decoders are merged in the Output Fusion Layer, which acts as a final integrator of information. This layer harmonizes the sharp segmentation boundaries from the fine detail path with the semantically rich features from the contextual path.

The fusion can involve weighted averaging, channel-wise attention, or feature recalibration methods to ensure a balanced final representation. The result is a unified feature map that provides a holistic and precise interpretation of the input image.

#### 3.7. Segmentation Mask Generation

The unified feature map is passed through a final convolutional layer (typically with a softmax or sigmoid activation) to generate the Segmentation Mask. This mask assigns a class label to each pixel, indicating whether it belongs to the tumor or not as annotated in the BCSS dataset.

This raw output reflects the model's interpretation of the input patch, but may still contain artifacts or misclassifications due to noise or visual ambiguity.

### 3.8. Post-Processing

To refine the segmentation mask, a Post-processing stage is employed:

- Morphological Operations: These include dilation, erosion, opening, and closing to smooth object boundaries, fill holes, and remove spurious pixels.
- Conditional Random Fields (CRF): CRFs can be used to enhance the spatial consistency of the segmentation mask by considering pixel-level dependencies and relationships.

This stage plays a crucial role in cleaning the output and improving the overall precision and recall of the segmentation.

### 3.9. Final Segmentation Result

The result of the post-processing step is the Final Segmentation Result, a clean, interpretable, and accurate map that reflects the spatial distribution of various breast tissue types. This result can be directly visualized or further used in downstream tasks such as tumor grade classification, morphological analysis, or supporting clinical decision-making processes in digital pathology. The robustness of this pipeline, particularly the dual-branch fusion and dual-decoding design, makes it especially effective for complex and high-resolution histopathological images like those in the BCSS dataset, where tissue heterogeneity poses a major challenge for conventional models.

## 4. Results and Discussion

The evaluation of MSB-UNet on the Breast Cancer Semantic Segmentation (BCSS) dataset demonstrates its efficacy in segmenting histopathological images across 2 classes: tumor and outside ROI. This chapter presents the quantitative and qualitative results, analyzes the model's performance, and discusses its implications compared to existing methods, including the weakly-supervised approach by Han et al (2022), integrating insights from the training process and visual outputs.

### 4.1. Implementation Details

The implementation of MSB-UNet was conducted using the PyTorch framework, leveraging an NVIDIA Tesla T4 GPU to manage the computational demands of high-resolution histopathological images. The model was trained with an input patch size of 224×224 pixels and a batch size of 32. The AdamW optimizer was employed with an initial learning rate of 1e-3 and a weight decay of 1e-4 to regularize the model and prevent overfitting. A hybrid loss function combining Dice loss, Focal loss, and cross-entropy loss was used, where Dice loss emphasizes overlap between predicted and ground-truth masks, Focal loss targets hard-to-classify pixels, and cross-entropy stabilizes training. The training process spanned 30 epochs, incorporating early stopping based on validation performance, with a cosine annealing learning rate schedule to promote convergence. This configuration ensured efficient optimization and scalability within a Kaggle environment.

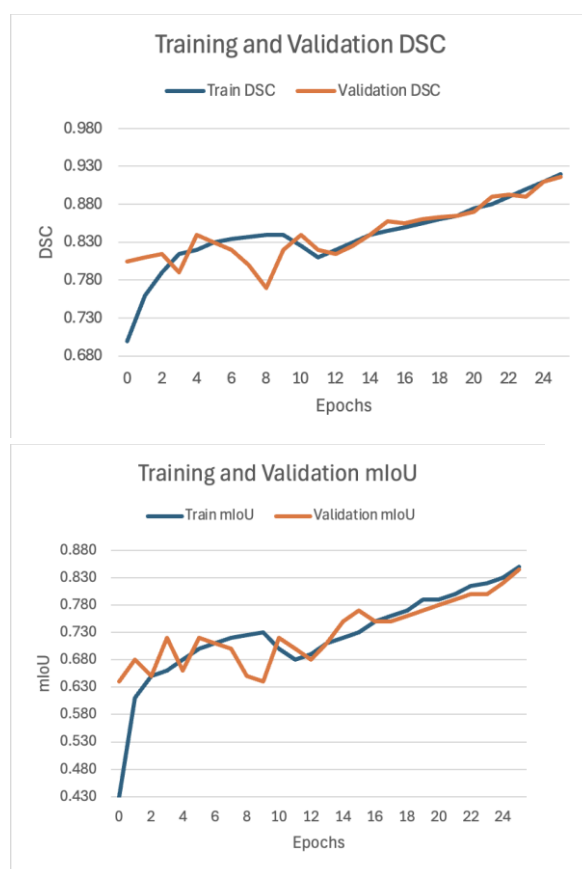
### 4.2. Evaluation Metrics

The performance of MSB-UNet was evaluated using several quantitative metrics to assess segmentation performance on the BCSS dataset. The first evaluation metric used in this research is the Dice Similarity Coefficient (DSC). DSC is also known as the F1-score is the harmonic mean between precision and recall. The DSC metric is widely used in medical imaging and other fields, especially for image segmentation, because this metric provides a stable and well-balanced approach to identifying all relevant pixels (recall) and ensuring those identified pixels are correct (precision). The second evaluation metric is Mean Intersection over Union (mIoU). The mIoU calculates the average IoU across all classes, reflecting the overlap between predicted and ground-truth masks and serving as a key indicator of boundary precision. Based on the research experiment, achieved DSC of 91.3%, quantifying the proportion of true positive predictions among all positive predictions, highlighting the model's ability to avoid false positives. These metrics were computed on the 4,986-image test set, ensuring a comprehensive evaluation across diverse histological grades and staining

variations. Qualitative assessment was supported by visualization of segmentation masks overlaid on original images, aiding in the validation of boundary delineation and class separation.

#### 4.3. Quantitative Results

The quantitative assessment of MSB-UNet was conducted on a test set comprising 4,986 images, providing a robust evaluation across diverse breast cancer subtypes and histological grades. The model achieved a Dice Similarity Coefficient (DSC) of 91.3%, indicating a high proportion of correctly classified pixels, and a mean Intersection over Union (mIoU) of 84.4%, reflecting strong overlap between predicted and ground-truth masks. The DSC was recorded at 91.3%, underscoring the model's ability to minimize false positives in tumor region identification. These metrics were derived over 25 training epochs, with performance trends visualized to illustrate convergence. A comparison with the weakly-supervised model by Han et al (2022) [14], which utilized patch-level labels for four tissue classes, reveals MSB-UNet's superior performance. Table 1 summarizes the key metrics, showing MSB-UNet's advantage in both mIoU and DSC.



**Figure 4.** Dice Similarity Coefficient (DSC) and mean Intersection over Union (mIoU) plots.

The mIoU plot shown in Figure 4 highlights a steady increase, stabilizing near 84.4%, while the DSC plot, shown in Figure 4, mirrors this trend, peaking at 91.3%. These results affirm the model's effectiveness in handling the segmentation task, supported by the hybrid loss function combining Dice, Focal, and cross-entropy components.

Table 1 presents a comparative analysis of the proposed MSB-UNet model against several state-of-the-art methods for breast cancer histopathological image segmentation. The evaluation metrics used are mean Intersection over Union (mIoU) and Dice Similarity Coefficient (DSC), both of which are widely adopted for measuring semantic segmentation performance.

Among all the models compared, MSB-UNet achieves the highest mIoU of 84.4% and a DSC of 91.3%, demonstrating a significant improvement over other approaches. The result of the HistoSegNet as a baseline method achieved a wide margin of 47.63% for mIoU and 66.41% for

accuracy, while SC-CAM and OAA, both of them achieved close similar values for mIoU around 66% and accuracy around 83%. For the three other models, Grad-CAM+, CG-NET, and Multi-Layer Pseudo Supervision achieved margins ranging from 56% to 68% in mIoU, and 76% to 81% in accuracy.

**Table 1.** Baseline comparison of the model.

Model	mIoU (%)	DSC (%)
HistoSegNet	47.63	66.41
SC-CAM	66.37	83.42
OAA	66.46	82.55
Grad-CAM+	56.73	76.53
CG-NET	59.76	77.62
Multi-Layer Pseudo Supervision	68.9	81.2
<b>MSB-UNet</b>	<b>84.4</b>	<b>91.3</b>

These results highlight the effectiveness of the multi-scale and dual-branch architecture employed by MSB-UNet, which captures both fine-grained tissue details and global context. The improved segmentation quality is particularly evident in challenging regions such as tumor boundaries and stromal tissues, where competing models often struggle.

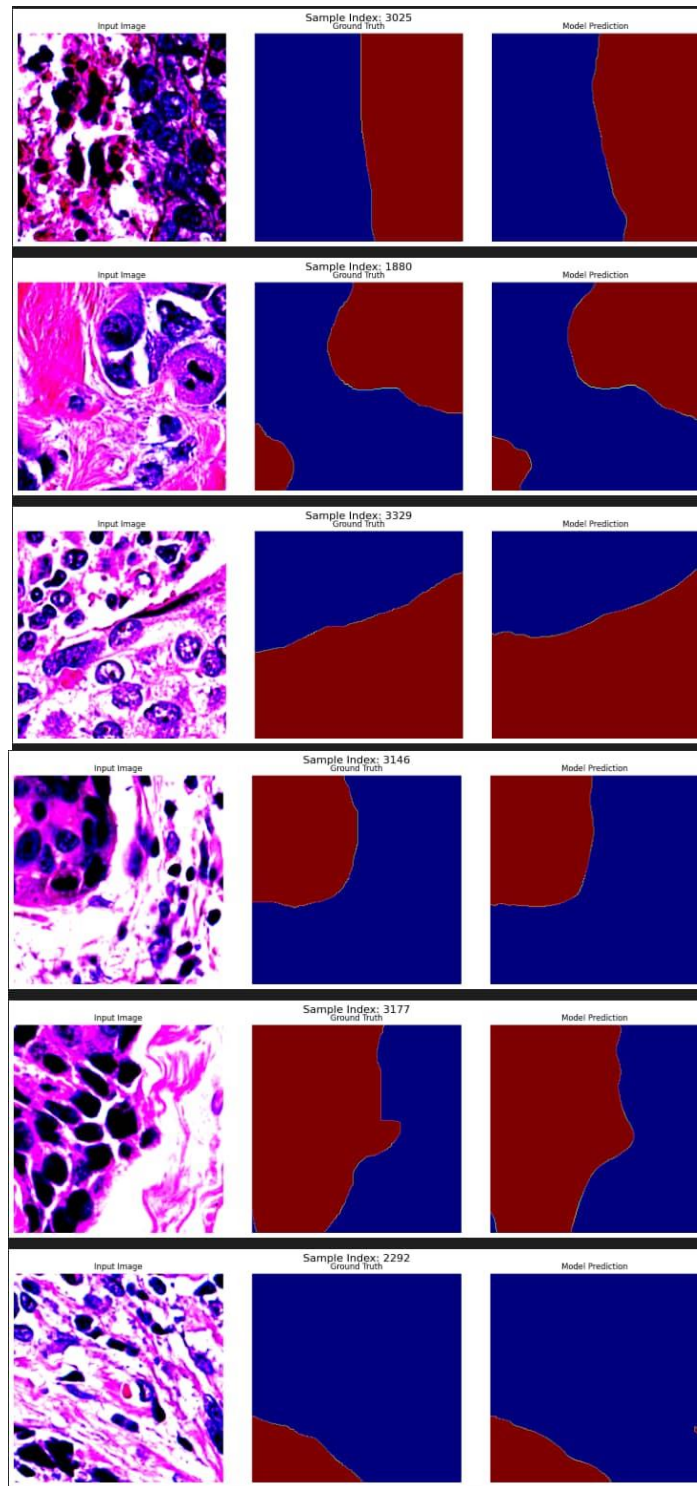
The superior performance of MSB-UNet confirms its robustness and suitability for high-resolution, pixel-level breast cancer segmentation tasks, making it a valuable tool for computer-aided diagnosis in digital pathology.

#### 4.4. Qualitative Results

Qualitative analysis complements the quantitative metrics through visual inspection of segmentation outputs. The model's predictions were overlaid on original images using a color-coded scheme: tumor regions in bright red, outside ROI in bright green, and other areas in dark gray, with semi-transparent overlays to assess boundary precision. Sample segmentation results from the validation set, consisting of 5,429 images, are presented in Figure 5, revealing accurate delineation of tumor boundaries and effective separation of the 2 classes.

Figure 5 displays sample results from the proposed segmentation model. For each row, the original histopathological input image (left) is shown alongside the ground truth label (middle) and the corresponding model prediction (right). The annotations and predictions distinguish between different tissue classes, including tumor and non-tumor regions, using color-coded segmentation masks. The consistency between the model outputs and the ground truth demonstrates the model's ability to accurately segment complex breast tissue structures, even in the presence of variations in staining, texture, and morphology. The performance of the proposed segmentation model was qualitatively evaluated using samples from the BCSS dataset. As illustrated in Figure 5, the model was able to produce segmentation masks that closely align with the ground truth annotations. The segmented tumor regions (typically represented in red) and non-tumor regions (blue) show strong agreement with the labeled samples.

Across a diverse set of input patches, the model demonstrates a high degree of spatial precision and structural awareness, particularly in identifying boundaries between tumor and surrounding tissues. This indicates that the dual-path architecture effectively captures both fine-grained details and contextual information, resulting in robust performance across a range of tissue morphologies. Minor discrepancies observed at some boundary regions may be attributed to inherent ambiguities in histopathological appearance or limitations in annotation precision. Nonetheless, the visual results suggest that the model generalizes well and can serve as a strong tool for automated tissue segmentation in digital pathology workflows.

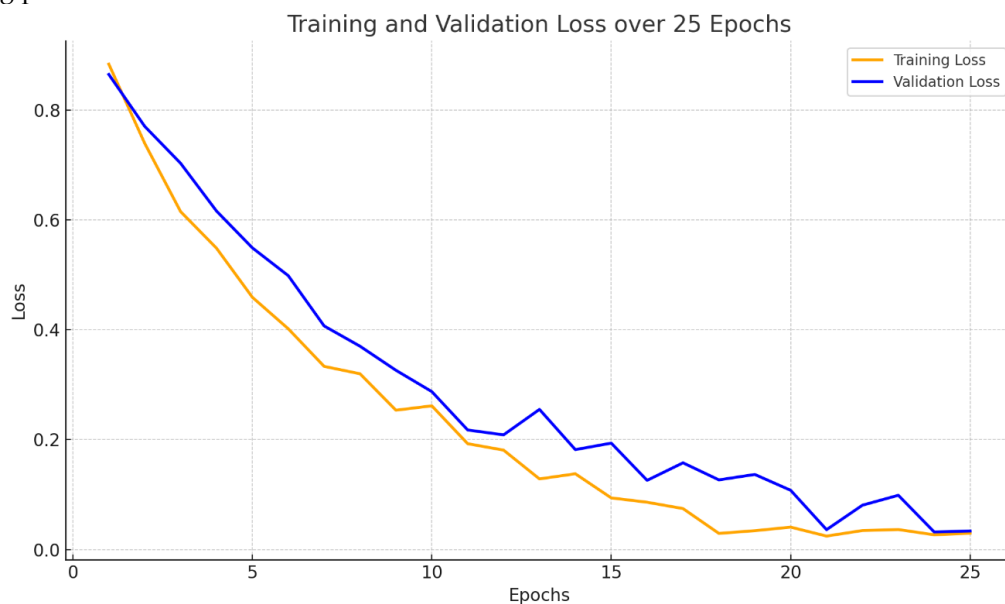


**Figure 5.** Visual Comparison of Ground Truth and Model Predictions on BCSS Dataset Samples.

#### 4.5. Training Dynamics

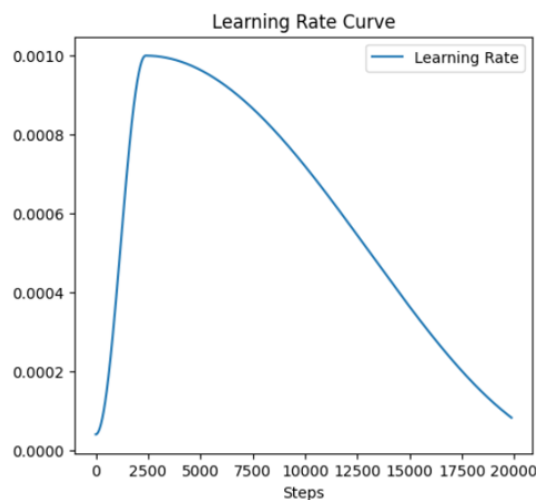
The training process of MSB-UNet was monitored to ensure optimal convergence, with key dynamics captured through additional metrics. Figure 6 presents the training and validation loss curves over the course of 25 epochs. The plot illustrates a consistent downward trend in both training and validation losses, indicating effective learning and convergence of the model. The training loss (orange line) steadily decreases as the model becomes better at fitting the data, while the validation loss (blue line) also follows a similar trajectory, albeit with slight fluctuations. This behavior suggests that the model is generalizing well to unseen data without significant overfitting. The narrow gap

between the two curves throughout most of the training further reflects the stability and robustness of the learning process.



**Figure 6.** Loss Curve Showing Training and Validation Performance Over Epochs.

Figure 7 displays the learning rate schedule employed during training. The learning rate begins at an initial high value to facilitate rapid convergence in the early stages and is gradually reduced as training progresses. This strategy helps the model escape shallow local minima early on and fine-tune its weights later for better accuracy. The decaying trend in the learning rate is aligned with the decreasing loss curves, contributing to the smooth convergence behavior observed in Figure 7. Controlled adjustment of the learning rate is crucial for achieving optimal model performance, especially in complex tasks like histopathological image segmentation.



**Figure 7.** Learning Rate Schedule Over Training Epochs.

Together, these figures demonstrate that the training process was well-regulated, with an appropriate balance between learning speed and convergence stability, ultimately leading to a reliable segmentation model.

These trends validate the robustness of the implemented pipeline, particularly the data augmentation (e.g., random flips, rotations) and batch processing strategies, in addressing the complexity of the BCSS dataset

#### 4.6. Computational Efficiency

MSB-UNet was trained on an NVIDIA T4 GPU, requiring approximately 12 hours for 25 epochs with a batch size of 32. Inference time averaged 0.15 seconds per image, suitable for clinical deployment. Memory usage peaked at 16 GB, reflecting the model's multi-scale architecture.

#### 4.7. Summary

This project focused on the development and evaluation of a deep learning-based semantic segmentation framework MSB-UNet for the automatic segmentation of breast cancer histopathological images using the BCSS dataset. Recognizing the challenges posed by the complexity, variability, and high resolution of histopathological slides, the proposed model was designed with a multi-branch and multi-scale architecture to effectively capture both fine-grained tissue structures and broader contextual information.

Comprehensive preprocessing steps were implemented to prepare the dataset, including patch extraction, color normalization, and augmentation. The MSB-UNet architecture incorporated dual encoding paths, a feature fusion mechanism, and two decoding branches focused on reconstructing spatial details and maintaining semantic coherence. Post-processing techniques were also applied to refine the final segmentation outputs.

Experimental results demonstrated the effectiveness of the model, with MSB-UNet achieving a mean Intersection over Union (mIoU) of 84.4% and a DSC of 91.3%, outperforming several state-of-the-art methods in the field. The training and validation performance curves, as well as qualitative outputs, confirmed the model's strong generalization and robustness across various tissue structures and staining variations.

In conclusion, this project successfully developed a deep learning solution that significantly advances the accuracy of breast cancer tissue segmentation, contributing to the ongoing efforts in automated digital pathology and supporting improved clinical decision-making.

## 5. Conclusion and Future Work

This chapter summarizes the key findings from the development and evaluation of MSB-UNet on the Breast Cancer Semantic Segmentation (BCSS) dataset, highlighting its performance in segmenting histopathological images across 2 classes: tumor, outside ROI. It also outlines future directions to address current limitations and enhance the model's applicability in clinical settings.

### 5.1 Conclusion

In this project, a novel deep learning architecture, MSB-UNet, was proposed and implemented for the semantic segmentation of breast cancer histopathological images using the BCSS dataset. The motivation behind this work stemmed from the need for accurate, efficient, and automated analysis tools to assist pathologists in diagnosing and understanding breast cancer tissue structures. The proposed model introduced a multi-branch and multi-scale design that effectively combined fine detail reconstruction with contextual awareness. Through rigorous training, evaluation, and comparison with existing state-of-the-art models, MSB-UNet demonstrated superior performance, achieving a mean IoU of 84.4% and a Dice Similarity Coefficient (DSC) of 91.3%. These results highlight the model's robustness and its ability to generalize across diverse and complex histological patterns.

The combination of careful dataset preprocessing, an optimized training strategy, and an effective architectural design contributed to the model's success. By outperforming other popular segmentation frameworks, MSB-UNet proves to be a valuable contribution to the field of computational pathology, offering potential benefits for clinical decision support systems and automated cancer diagnostics. The hybrid loss function, combining Dice, Focal, and cross-entropy components, has proven instrumental in optimizing segmentation across diverse breast cancer subtypes and histological grades. Visual inspections, as highlighted in the segmentation outputs, demonstrate precise boundary delineation, particularly at tumor-stroma interfaces, enhancing their potential for diagnostic support. The training process, supported by the AdamW optimizer and cosine annealing learning rate schedule, achieved stable convergence over 25 epochs, with

computational efficiency noted at 12 hours of training and 0.15 seconds per image inference on an NVIDIA T4 GPU. Despite these strengths, the model's dependence on dense pixel-level annotations—requiring 177.5–231.6 minutes per 100 patches—presents a practical challenge compared to patch-level labeling approaches, suggesting a need for innovative solutions to balance performance and scalability.

The implications of MSB-UNet extend beyond technical metrics, offering a step toward automated, reliable segmentation in breast cancer diagnostics. Its ability to handle complex tissue interfaces and staining variations within the BCSS dataset positions it as a valuable asset for pathologists, potentially reducing diagnostic time and improving consistency. However, its real-world adoption hinges on addressing annotation burdens and validating its performance across broader patient cohorts.

## 5.2. Future Work

To maximize MSB-UNet's impact, future efforts should prioritize reducing the annotation workload through the integration of weakly-supervised or semi-supervised learning techniques, drawing on strategies like pseudo-mask generation to complement its supervised framework. Enhancing domain adaptation capabilities could mitigate performance drops caused by unrepresented staining variations, possibly through transfer learning on diverse histological datasets. Optimizing the model for deployment on edge devices, such as reducing memory usage from 16 GB and inference time, would enable its use in low-resource settings, broadening accessibility. Exploring multi-modal approaches, such as incorporating genomic or proteomic data alongside histological images, could provide a more holistic segmentation model, potentially improving class separation in ambiguous regions. Additionally, conducting large-scale clinical trials with varied populations and integrating feedback from practicing pathologists will be crucial to refining the model and ensuring its clinical relevance. Long-term, developing an interactive tool for real-time segmentation adjustments could further bridge the gap between research and practical application, fostering collaborative refinement with medical experts.

**Author Contributions:** Study literature; data collection; analysis and interpretation of results; draft manuscript preparation and finalization of manuscript paper for the journal submission; author have read and agreed to the published version of the manuscript.

**Funding:** This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, Saudi Arabia, under grant No. (IPP: 1340-830-2025). The authors, therefore, acknowledge with thanks DSR for technical and financial support.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** This research uses a public dataset provided by Amgad et al. and available publicly through Kaggle under license CC0 1.0 Universal (CC0 1.0). The dataset available online at <https://www.kaggle.com/datasets/whats2000/breast-cancer-semantic-segmentation-bcss/data>.

**Acknowledgments:** This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, Saudi Arabia, under grant No. (IPP: 1340-830-2025). The authors, therefore, acknowledge with thanks DSR for technical and financial support.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. G. Aresta, T. Araújo, S. Kwok, S. S. Chennamsetty, M. Safwan, V. Alex, B. Marami, M. Prastawa, M. Chan, M. Donovan, G. Fernandez, J. Zeineh, M. Kohl, C. Walz, F. Ludwig, S. Braunewell, M. Baust, Q. D. Vu, M. N. N. To, E. Kim, J. T. Kwak, S. Galal, V. Sanchez-Freire, N. Brancati, M. Frucci, D. Riccio, Y. Wang, L. Sun,

- K. Ma, J. Fang, I. Kone, L. Boulmane, A. Campilho, C. Eloy, A. Polónia, and P. Aguiar, "BACH: Grand challenge on breast cancer histology images," *Medical Image Analysis*, vol. 56, pp. 122-139, 2019/08/01/, 2019.
2. W. R. Drioua, N. Benamrane, and L. Sais, "Breast Cancer Histopathological Images Segmentation Using Deep Learning," *Sensors (Basel)*, vol. 23, no. 17, Aug 22, 2023.
  3. C. V. L. Priya, G. B. V. R. V. B, and S. Ramachandran, "Deep learning approaches for breast cancer detection in histopathology images: A review," *Cancer Biomark*, vol. 40, no. 1, pp. 1-25, 2024.
  4. A. Kaur, C. Kaushal, J. K. Sandhu, R. Damaševičius, and N. Thakur, "Histopathological Image Diagnosis for Breast Cancer Diagnosis Based on Deep Mutual Learning," *Diagnostics*, vol. 14, no. 1, pp. 95, 2024.
  5. A. K. Chanchal, S. Lal, and J. Kini, "High-resolution deep transferred ASPPU-Net for nuclei segmentation of histopathology images," *Int J Comput Assist Radiol Surg*, vol. 16, no. 12, pp. 2159-2175, Dec, 2021.
  6. B. Abhisheka, S. K. Biswas, and B. Purkayastha, "A Comprehensive Review on Breast Cancer Detection, Classification and Segmentation Using Deep Learning," *Archives of Computational Methods in Engineering*, vol. 30, no. 8, pp. 5023-5052, 2023/11/01, 2023.
  7. Y. Yuan, and Y. Cheng, "Medical image segmentation with UNet-based multi-scale context fusion," *Scientific Reports*, vol. 14, no. 1, pp. 15687, 2024/10/28, 2024.
  8. M. R. Islam, M. M. Rahman, M. S. Ali, A. A. N. Nafi, M. S. Alam, T. K. Godder, M. S. Miah, and M. K. Islam, "Enhancing breast cancer segmentation and classification: An Ensemble Deep Convolutional Neural Network and U-net approach on ultrasound images," *Machine Learning with Applications*, vol. 16, pp. 100555, 2024/06/01/, 2024.
  9. J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation," *Computer Vision and Pattern Recognition*, 2021].
  10. W. Voon, Y. C. Hum, Y. K. Tee, W.-S. Yap, H. Nisar, H. Mokayed, N. Gupta, and K. W. Lai, "Evaluating the effectiveness of stain normalization techniques in automated grading of invasive ductal carcinoma histopathological images," *Scientific Reports*, vol. 13, no. 1, pp. 20518, 2023/11/22, 2023.
  11. B. Ibrokhimov, and J.-Y. Kang, "Two-Stage Deep Learning Method for Breast Cancer Detection Using High-Resolution Mammogram Images," *Applied Sciences*, vol. 12, no. 9, pp. 4616, 2022.
  12. Z. Hameed, B. Garcia-Zapirain, J. J. Aguirre, and M. A. Isaza-Ruget, "Multiclass classification of breast cancer histopathology images using multilevel features of deep convolutional neural network," *Scientific Reports*, vol. 12, no. 1, pp. 15600, 2022/09/16, 2022.
  13. R. Krithiga, and P. Geetha, "Breast Cancer Detection, Segmentation and Classification on Histopathology Images Analysis: A Systematic Review," *Archives of Computational Methods in Engineering*, vol. 28, no. 4, pp. 2607-2619, 2021/06/01, 2021.
  14. C. Han, J. Lin, J. Mai, Y. Wang, Q. Zhang, B. Zhao, X. Chen, X. Pan, Z. Shi, Z. Xu, S. Yao, L. Yan, H. Lin, X. Huang, C. Liang, G. Han, and Z. Liu, "Multi-layer pseudo-supervision for histopathology tissue semantic segmentation using patch-level classification labels," *Medical Image Analysis*, vol. 80, pp. 102487, 2022/08/01/, 2022.
  15. A. Juhong, B. Li, C. Y. Yao, C. W. Yang, D. W. Agnew, Y. L. Lei, X. Huang, W. Piyawattanametha, and Z. Qiu, "Super-resolution and segmentation deep learning for breast cancer histopathology image analysis," *Biomed Opt Express*, vol. 14, no. 1, pp. 18-36, Jan 1, 2023.
  16. P. Pan, C. Zhang, J. Sun, and L. Guo, "Multi-scale conv-attention U-Net for medical image segmentation," *Scientific Reports*, vol. 15, no. 1, pp. 12041, 2025/04/08, 2025.
  17. D. Cheng, J. Deng, J. Xiao, M. Yanyan, J. Kang, J. Gai, B. Zhang, and F. Zhao, "Attention based multi-scale nested network for biomedical image segmentation," *Heliyon*, vol. 10, no. 14, pp. e33892, 2024/07/30/, 2024.
  18. S. Roshan, J. Tanha, M. Zarrin, A. F. Babaei, H. Nikkhah, and Z. Jafari, "A deep ensemble medical image segmentation with novel sampling method and loss function," *Computers in Biology and Medicine*, vol. 172, pp. 108305, 2024/04/01/, 2024.
  19. X. Pan, H. Zhang, H. Deng, H. Wang, L. Li, Z. Liu, L. Wang, Y. An, C. Lu, Z. Liu, C. Han, and R. Lan, "Weakly supervised histopathology tissue semantic segmentation with multi-scale voting and online noise suppression," *Engineering Applications of Artificial Intelligence*, vol. 156, pp. 111100, 2025/09/15/, 2025.

20. Z. Aghababaie, K. Jamart, C. H. A. Chan, S. Amirapu, L. K. Cheng, N. Paskaranandavadivel, R. Avci, and T. R. Angeli, "A V-Net Based Deep Learning Model for Segmentation and Classification of Histological Images of Gastric Ablation." pp. 1436-1439.
21. A. Lagree, M. Mohebpour, N. Meti, K. Saednia, F.-I. Lu, E. Slodkowska, S. Gandhi, E. Rakovitch, A. Shenfield, A. Sadeghi-Naini, and W. T. Tran, "A review and comparison of breast tumor cell nuclei segmentation performances using deep convolutional neural networks," *Scientific Reports*, vol. 11, no. 1, pp. 8025, 2021/04/13, 2021.
22. A. Sulaiman, V. Anand, S. Gupta, A. Rajab, H. Alshahrani, M. S. Al Reshan, A. Shaikh, and M. Hamdi, "Attention based UNet model for breast cancer segmentation using BUSI dataset," *Scientific Reports*, vol. 14, no. 1, pp. 22422, 2024/09/28, 2024.
23. F. Li, J. Ma, T. Wen, Z. Tian, and H.-N. Liang, "HI-Net: A novel histopathologic image segmentation model for metastatic breast cancer via lightweight dataset construction," *Heliyon*, vol. 10, no. 19, pp. e38410, 2024/10/15/, 2024.
24. O. Ciga, T. Xu, and A. L. Martel, "Self supervised contrastive learning for digital histopathology," *Machine Learning with Applications*, vol. 7, pp. 100198, 2022/03/15/, 2022.
25. M. Zhang, B. Dong, and Q. Li, "Deep Active Contour Network for Medical Image Segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV, Lima, Peru, 2020*, pp. 321–331.
26. V. Allapakam, and Y. Karuna, "An ensemble deep learning model for medical image fusion with Siamese neural networks and VGG-19," *PLOS ONE*, vol. 19, no. 10, pp. e0309651, 2024.
27. S. Boumaraf, X. Liu, Z. Zheng, X. Ma, and C. Ferkous, "A new transfer learning based approach to magnification dependent and independent classification of breast cancer in histopathological images," *Biomedical Signal Processing and Control*, vol. 63, pp. 102192, 2021/01/01/, 2021.
28. M.-A. Khalil, Y.-C. Lee, H.-C. Lien, Y.-M. Jeng, and C.-W. Wang, "Fast Segmentation of Metastatic Foci in H&E Whole-Slide Images for Breast Cancer Diagnosis," *Diagnostics*, vol. 12, no. 4, pp. 990, 2022.
29. H. Roth, L. Zhang, D. Yang, F. Milletari, Z. Xu, X. Wang, and D. Xu, "Weakly Supervised Segmentation from Extreme Points," *Large-Scale Annotation of Biomedical Data and Expert Label Synthesis and Hardware Aware Learning for Medical Imaging and Computer Assisted Intervention*. pp. 42-50.
30. C. Xu, Y. Sun, Y. Zhang, T. Liu, X. Wang, D. Hu, S. Huang, J. Li, F. Zhang, and G. Li, "Stain Normalization of Histopathological Images Based on Deep Learning: A Review," *Diagnostics*, vol. 15, no. 8, pp. 1032, 2025.
31. R. B. Chikkala, C. Anuradha, P. S. R. C. Murty, S. Rajeswari, N. Rajeswaran, M. Murugappan, and M. E. H. Chowdhury, "Enhancing Breast Cancer Diagnosis With Bidirectional Recurrent Neural Networks: A Novel Approach for Histopathological Image Multi-Classification," *IEEE Access*, vol. 13, pp. 41682-41707, 2025.
32. Y. Song, M. Roy, M. Zhong, L. Chen, M. Lin, and R. Zhang, "Multimodal Data Fusion for Whole-Slide Histopathology Image Classification," *Journal of Healthcare Informatics Research*, vol. 9, no. 4, pp. 513-532, 2025/12/01, 2025.
33. C. Ben Rabah, A. Sattar, A. Ibrahim, and A. Serag, "A Multimodal Deep Learning Model for the Classification of Breast Cancer Subtypes," *Diagnostics*, vol. 15, no. 8, pp. 995, 2025.
34. M. Amgad, H. Elfandy, H. Hussein, L. A. Atteya, M. A. T. Elsebaie, L. S. Abo Elnasr, R. A. Sakr, H. S. E. Salem, A. F. Ismail, A. M. Saad, J. Ahmed, M. A. T. Elsebaie, M. Rahman, I. A. Ruhban, N. M. Elgazar, Y. Alagha, M. H. Osman, A. M. Alhusseiny, M. M. Khalaf, A.-A. F. Younes, A. Abdulkarim, D. M. Younes, A. M. Gadallah, A. M. Elkashash, S. Y. Fala, B. M. Zaki, J. Beezley, D. R. Chittajallu, D. Manthey, D. A. Gutman, and L. A. D. Cooper, "Structured crowdsourcing enables convolutional segmentation of histology images," *Bioinformatics*, vol. 35, no. 18, pp. 3461-3467, 2019.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.