STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–20.
Published online in International Academic Press (www.IAPress.org)

CancerSeg-XA: Medical Histopathology Segmentation System Based on Xception Backbone and Attention Mechanisms

Alaa Mohamed Youssef [1,*], Wessam Hassan El Behaidy [1,2], Aliaa Abdel-Haleim Abdel-Razik Youssif [3]

[1] *Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt*
[2] *Faculty of Informatics and Computer Science, British University in Egypt (BUE), El-Sherouk, Egypt*
[3] *Arab Academy for Science, Technology and Maritime Transport (AASTMT), Cairo, Egypt*

Abstract Accurate segmentation of histopathological images is essential to support early diagnosis and effective treatment planning in cancer care. This study presents CancerSeg-XA, a deep learning-based histopathology segmentation system designed to deliver robust performance across diverse tissue types and imaging sources. Built upon the DeepLabV3+ framework, CancerSeg-XA incorporates architectural enhancements to strengthen feature representation and improve model stability. The system was evaluated on three widely recognized datasets—BCSS, PanNuke, and PUMA—each presenting distinct structural and clinical challenges. Across all datasets, CancerSeg-XA consistently outperformed the baseline DeepLabV3+ in terms of segmentation accuracy, recall, and F1-score. Specifically, it achieved accuracy improvements of 4.78%, 4.31%, and 3.22% on BCSS, PanNuke, and PUMA, respectively, along with substantial gains in FwIoU. These results highlight the model's ability to generalize effectively across varied histopathological contexts, positioning CancerSeg-XA as a promising solution for clinical integration and future research in digital pathology.

Keywords Deep learning, cancer segmentation, Histopathological images, DeepLabV3+, Xception backbone, Attention mechanism,PanNuke, PUMA

DOI: 10.19139/soic-2310-5070-3023

1. Introduction

Cancer remains one of the most pressing global health challenges, accounting for nearly 10 million deaths in 2020 [1]. This devastating disease arises from the transformation of normal cells into malfunctioning ones, which proliferate uncontrollably and invade the surrounding tissues. Cancer continues to pose one of the most formidable challenges of the 21st century, remaining among the leading causes of death globally, with an estimated 10 million deaths in 2020 alone, and its burden is projected to rise dramatically unless urgent action is taken [2]. Early detection and proper management are critical for the successful treatment and cure of many forms of cancer. However, the diagnostic process is often labor-intensive, requiring pathologists to examine the stained tissue samples under a microscope meticulously. The emergence of whole-slide imaging (WSI) and digital scanners has revolutionized this process, enabling the digitization of histopathological images for analysis at various magnifications [3].

This technological advancement has paved the way for computer vision to play a pivotal role in automating WSI analysis, significantly reducing diagnosis time and enhancing efficiency.

A key component of this automation is histopathological image segmentation, which involves the precise delineation of structures and regions within WSIs. Despite its importance, generating sufficient manually labeled

*Correspondence to: Alaa M. Youssef (Email: ayoussef2002@hotmail.com). Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

data to train convolutional neural network (CNN) models remains challenging. The annotation of gigapixel WSIs is time-consuming, costly, and requires specialized clinical expertise to account for the morphological heterogeneity of malignant tumors. Even among experts, variations in annotation practices, such as the use of different magnification levels to capture fine details, further complicate the process [4]. These challenges highlight the need for innovative approaches to alleviate the annotation burden and enhance the scalability of histopathological image segmentation.

Deep learning has made remarkable strides in medical diagnostics, particularly in the development of computer-aided diagnostic (CAD) systems. These systems leverage deep learning to help practitioners detect and differentiate abnormalities with greater precision [5]. The performance of CAD systems heavily depends on the selection of relevant features and the ability of the classifier to learn from these features. CNNs, in particular, have emerged as powerful tools for this purpose owing to their ability to learn robust feature representations and classify image data effectively. The success of CNNs depends on both the input representation of images and network architecture [6]. By enhancing the feature space, CNNs can achieve superior performance in classification tasks, making them well-suited for challenges such as mitosis detection and cancer-level classification in cancer segmentation [7].

Building on our previous work [8], which focused on breast cancer histopathology images, this study broadens the evaluation to multiple cancer types to assess the generalizability and stability of the proposed segmentation framework across diverse histopathological settings. In this extended design, the standard DeepLabV3+ architecture was enhanced through two key modifications: the integration of a spatial attention layer positioned between the encoder and decoder, and the replacement of the original ResNet50 backbone with the Xception backbone. These combined architectural improvements yielded the best overall performance, enabling CancerSeg-XA to capture finer structural details, enhance feature discrimination, and achieve superior segmentation accuracy across all evaluated datasets.

First, a dedicated spatial attention module is explicitly positioned between the encoder and decoder, processing the encoder's output before it is passed to the decoder. This targeted placement enables the module to refine spatial feature representations at a critical transition point, enhancing the model's ability to emphasize diagnostically relevant structures in complex histopathological regions. As a result, the integration of this attention mechanism improves both the stability and accuracy of the segmentation process. Second, we replaced the conventional ResNet50 backbone with the more efficient Xception architecture, which not only enhanced the feature extraction capabilities but also reduced the model complexity by lowering the number of trainable parameters and the overall computational costs. A comprehensive performance analysis was conducted, including the evaluation of training and validation loss, learning stability, and accuracy, all of which demonstrated significant improvements over the baseline DeepLabV3+ model and reinforced the effectiveness of the proposed extension.

This study presents a comprehensive medical histopathology segmentation system designed to operate effectively across diverse tissue types and under different imaging conditions. By enhancing the standard DeepLabV3+ architecture with targeted modifications, the proposed system aims to improve segmentation accuracy, learning stability, and computational efficiency. The key contributions of this study are as follows:

- We propose an enhanced DeepLabV3+ architecture by integrating a spatial attention mechanism between the encoder output and the decoder input, enabling the model to emphasize diagnostically relevant features better and to improve segmentation performance.
- The model adopts the Xception backbone instead of the traditional ResNet50, providing superior feature extraction capabilities while significantly reducing the number of trainable parameters and the computational cost.
- A thorough performance evaluation was conducted across three diverse histopathology datasets representing different organs and tissue characteristics. This analysis demonstrates the system's ability to generalize and maintain consistent accuracy across various medical imaging challenges.
- The system achieved notable improvements in accuracy, recall, and F1-score compared with the baseline DeepLabV3+, highlighting its potential for broader applicability in digital pathology and clinical workflows.

The remainder of this paper is organized as follows: Section II reviews prior research on cancer segmentation and detection. Section III details the dataset preparation, proposed model architecture, and validation measures.

Section IV presents and discusses the experimental results of this study. Section V concludes the paper with a summary of the findings.

2. Related work

Cancer segmentation using histopathology images has gained significant attention because of its critical role in accurate diagnosis and treatment planning of cancer. Histopathology, the microscopic examination of tissues, provides detailed insights into cellular morphology and tissue architecture, which are essential for identifying cancerous regions in tissues. To achieve this, a quantitative representation of tumor cells or tissues is often required, relying on the extraction of relevant features to accurately assess tissue and organ functions [9]. Feature extraction, which reduces an image to a compressed feature vector, plays a pivotal role in distinguishing between normal and malignant tissues and grading them appropriately [10].

Traditionally, feature extraction has involved handcrafted features, such as shape, color, and texture, which are particularly relevant to cancer histopathology. Key morphometric features of cells, such as the nuclear area, convex area, and cell outline, are often more pronounced in cancer cells than in normal cells. These cells exhibited dark nucleoli, reduced cytoplasm, disorganized chromosomes, abnormal growth patterns, and variations in shape and size. Studies have shown that using the mean values of these features for cell and nuclear outlines can significantly improve the grading process [11].

In recent years, deep learning architectures, particularly DeepLab, have been widely adopted and modified for various segmentation tasks, demonstrating remarkable improvements in accuracy and efficiency. For instance, DeepLabV3 has been integrated into pipelines for segmenting diverse histopathological structures, enhancing adaptability to different tissue types, and improving the overall segmentation quality [12]. Another study introduced an uncertainty-driven pooling layer into DeepLabV3+, which improved the segmentation accuracy and reliability by better handling ambiguous regions in histology images . [13].

DeepLabV3+ has also been applied in medical imaging, beyond histopathology. For example, it was used for liver tumor segmentation from CT images, combining deep convolutional neural networks (DCNNs) with probabilistic graphical models (DenseCRFs) to capture multiscale context information and refine target boundaries [14]. In cancer histopathology, DeepLabV3+ outperforms traditional methods, such as VGG16, ResNet18, SqueezeNet, and MobileNetV2, showcasing its superior segmentation capabilities [15].

The versatility of DeepLabV3+ extends to other domains. For instance, it has been employed for the semantic segmentation of buildings in high-resolution remote sensing images, leveraging its ability to handle high-resolution inputs and effectively capture multi-scale features [16]. In microscopic cancer segmentation, DeepLabV3+ demonstrates exceptional accuracy in classifying and segmenting cancerous regions [17]. Similarly, it achieves high accuracy and robustness in segmenting smooth muscle fibers in hematoxylin and eosin (H&E)-stained images [18] and cancerous tissues, owing to its encoder-decoder structure that enables detailed and precise segmentation [19].

In rectal cancer, DeepLabV3+ with an Xception-65 backbone was used to recognize pathological T-stages and tumor invasion, employing atrous spatial pyramid pooling for multiscale feature extraction [20, 21]. Beyond medical imaging, DeepLabV3+ has been applied to segment cracks on concrete surfaces, outperforming architectures such as U-Net and fully convolutional networks (FCN) [22]. It also surpassed UNet++ in accuracy for oral epithelial dysplasia (OED) segmentation [23].

Han et al. [4] applied DeepLabV3+ to the ”Breast Cancer Semantic Segmentation” (BCSS) dataset [24], introducing a modified ASPP module with varying atrous rates and a hybrid loss function combining cross-entropy and Dice loss. These enhancements led to improved segmentation results, particularly for small and irregularly shaped regions, with an accuracy of 84.83%.

Srijay et al. [25] proposed an advanced breast cancer tissue segmentation framework that incorporates spatially adaptive normalization layers into a ResNet-based architecture. These layers help retain fine-grained details that are typically lost in deep learning segmentation processes, thus improving segmentation accuracy. The model processes histopathological image tiles and generates refined segmentation outputs, achieving an overall accuracy of 77%

on the BCSS dataset. Huang et al. [26] developed a two-stage weakly supervised semantic segmentation model that focuses on capturing inter-class relationships among tissue types in histopathology images. Their method integrates a graph-parsing attention mechanism to enhance pseudo-label refinement and leverages a cyclic pseudo-mask strategy along with multi-resolution supervision. These innovations collectively improved the segmentation precision, resulting in an accuracy of 85.94% on the BCSS dataset.

Mauricio et al. [27]introduced an enhanced version of the U-Net architecture specifically designed for multiclass segmentation of breast cancer tissues. The model incorporates densely connected blocks and residual connections, which strengthen feature extraction and representation. This design enabled the model to effectively manage the complex and heterogeneous nature of breast cancer histopathological data, yielding an impressive segmentation accuracy of 81% on the BCSS dataset used.

The PanNuke dataset was originally introduced for nucleus instance segmentation, where each nucleus is delineated separately and assigned one of several nucleus classes. A substantial portion of the literature has therefore focused on improving instance-level prediction on PanNuke. Im et al. [28] proposed HDA-Net, a dual-attention architecture that leverages both RGB information and its color-decomposed HER representation to enhance nuclei localization and shape prediction through a multitask decoder. Their method achieved notable gains in Aggregated Jaccard Index (AJI) and Dice score, improving performance by 0.7Pina et al. [29] introduced CellNuc-DETR, which shifts nuclei analysis from pixel-wise segmentation to direct detection using a DETR-based hierarchical encoder with multi-scale deformable attention. Their approach achieved state-of-the-art detection and multi-class classification performance on PanNuke while delivering significantly faster inference on whole-slide images. Zhao et al. [30] developed SBC-UNet3+, a multi-branch UNet3+ variant incorporating segmentation, boundary detection, and classification-guided modules. The model demonstrated strong performance on PanNuke with 89.61% accuracy, 89.19% F1-score, 99.08% Dice, and 83.16% mIoU, outperforming HoverNet and other prior models. Ramesh et al. [31] proposed MANS-Net, which integrates spatial and channel attention to enhance nuclei delineation across diverse organs; the model achieved 91.85% accuracy and an F1-score of 89.97% on PanNuke. Chen et al. [32] introduced CellViT, a hybrid CNN–Vision Transformer framework that captures both local texture and long-range dependencies, reporting competitive results with 87.5% F1-score on PanNuke.

The following works adapted the PanNuke dataset for semantic segmentation by merging nucleus instances into the same label, treating all nuclei as a single class for binary segmentation to focus on overall nuclei delineation rather than individual instance separation.

Arnob et al. [33] introduced an extension to the Segment Anything Model (SAM) for pathological primitive segmentation, using a pre-trained SAM encoder for feature extraction, followed by cascaded feature propagation layers for bounding box localization and classification. The approach generates bounding boxes as zero-shot prompts for SAM's mask decoder, enabling end-to-end detection and segmentation without additional training. Tested on the PanNuke dataset (adapted for binary semantic segmentation), it achieved state-of-the-art F1 scores for nuclei detection (up to 85%) and binary panoptic quality (bPQ) for segmentation, while offering faster inference and cross-domain generalization to datasets like CoNSeP and MoNuSeg.

Koganti et al. [34] presented standardized workflows for reorganizing the PanNuke and NuCLS datasets, including training-validation splits, cell-type-specific subsets, multi-scale patch generation, and synthetic data creation using StyleGAN2-ADA to address class imbalance and data scarcity. The reorganized datasets support semantic segmentation and object detection tasks with consistent annotations. Experiments on the adapted PanNuke dataset (for binary semantic segmentation) showed improved model performance in binary classification and multi-task evaluation, with synthetic data enhancing generalization, achieving up to 85% Dice scores in nuclei segmentation tasks.

Ding et al. [35] developed SNOW, a large-scale synthetic pathological dataset for breast cancer nuclei segmentation, generated using a StyleGAN-based image synthesizer and HoVer-Net for automated annotations. The workflow produces 20k image tiles with 1,448,522 annotated nuclei, enabling supervised and semi-supervised training. While primarily focused on breast cancer, the method was extended to PanNuke-like scenarios for binary semantic segmentation, where models trained on SNOW achieved competitive Dice scores (around 0.85) compared to real-world data, boosting performance in low-data regimes by up to 5%.

Khaled et al. [36] proposed a semantic segmentation method using Segformer-b0 and Segformer-b4 architectures, which combine hierarchical Transformer encoders with lightweight MLP decoders for efficient multi-scale feature extraction without complex decoders. Evaluated on the PanNuke dataset (adapted for binary semantic segmentation), Segformer-b4 achieved a state-of-the-art mean Dice score of 0.845, outperforming RCSAU-Net (0.8482), while Segformer-b0 reached 0.82, demonstrating high efficiency and accuracy in histopathology nuclei segmentation.

Xia et al. [37] introduced AttE-Unet, a lightweight neural network for cell segmentation based on attention enhancement, incorporating attention mechanisms to boost feature extraction and combining deep learning with traditional image filtering for reduced computational demands. The model significantly cuts parameters and FLOPs to 1.57% and 0.1% of the original UNet. On the PanNuke dataset (adapted for binary semantic segmentation), AttE-Unet maintained 91.7% of the F1 score and 89.3% of the IoU compared to the baseline, with deployment on an MCU using only 2.09 MB Flash and 1.38 MB RAM, highlighting its suitability for edge devices.

Torbati et al. [38] proposed a multi-stage auto-context deep learning framework for simultaneous tissue and nuclei segmentation and classification in H&E-stained melanoma histology, using the PUMA dataset. Their pipeline consists of four stages: first, classifying input images as primary or metastatic; second, performing initial tissue segmentation; third, nuclei instance segmentation and classification; and finally, refining tissue segmentation using nuclei outputs and post-processing heuristics. Their approach secured first place in tissue segmentation and second in nuclei classification in the PUMA Challenge, achieving a micro-average tissue Dice score of 73.40% and a summed nuclei F1-score of 63.48% across the two challenge tracks.

In conclusion, DeepLabV3+ has demonstrated robust performance in various domains, including biomedical imaging, histopathology, MRI analysis, and remote sensing. With the targeted architectural modifications introduced in this study, DeepLabV3+ is a particularly promising framework for advancing histopathological image segmentation, offering improved generalization and accuracy across diverse datasets.

3. Materials and methods

This section discusses the dataset preparation, evaluation metrics, and design of the proposed system.

3.1. Dataset preparation

Three diverse histopathology datasets were used to comprehensively evaluate the proposed segmentation system. These datasets differ in terms of organ origin, annotation specificity, and tissue morphology, providing a robust benchmark for assessing the generalizability of the model across a wide spectrum of medical imaging scenarios. Each dataset underwent standardized preprocessing, data augmentation, and partitioning to align with the training and evaluation pipelines of the system. This multi-dataset strategy ensured that the model was rigorously tested not only for domain-specific performance but also for its adaptability to heterogeneous histopathological conditions.

Breast Cancer Semantic Segmentation (BCSS) dataset [24]. The dataset comprised 155 H&E-stained whole-slide images, with over 20,000 manually annotated tissue regions spanning 22 histological classes. To build a robust training set, we extracted 57,000 subimages representing tumor-rich regions of interest (ROIs). Each sub-image was 256 × 256 pixels in size and sampled using a 50% overlap strategy to maximize coverage. Data augmentation techniques were applied to enhance generalization. Additionally, sub-images containing classes representing less than 1% of the dataset were excluded to mitigate class imbalance and reduce label noise during training. The dataset was randomly partitioned into 42,800 images for training, 10,700 for validation, and 3,500 for testing.

The second dataset, PanNuke [39], offers a multi-organ histopathological benchmark comprising H&E images collected from various organs, including the breast, liver, kidney, prostate, and bladder. Each image was labeled with five cellular classes: neoplastic, inflammatory, connective tissue, epithelial, and dead cells. The original dataset contained 7,901 images; each was resized to 256 × 256 pixels for compatibility with the model. Data augmentation was employed to increase diversity, resulting in an expanded dataset of 47,406 images. These were split into 34132, 3792, and 9,482 for training, validation, and testing, respectively. The presence of multiple tissue origins makes PanNuke well-suited for testing the cross-domain generalizability of segmentation models.

To further assess the generalization of the proposed model across broader cancer types, we used the recently introduced Pan-cancer Unified Multicellular Annotation (PUMA) dataset [40]. This dataset comprised 103 high-resolution H&E-stained images, each with a resolution of 1024 × 1024 pixels. The ground-truth annotations were provided in the GeoJSON format, which we converted into the corresponding binary mask images of the same resolution for use in semantic segmentation. From these masks, 5,025 subimages of size 256 × 256 pixels were initially extracted to capture localized tissue patterns. These subimages were then expanded through data augmentation to a total of 28,332 images. The augmented dataset was subsequently divided into 21,708 training images, 5,424 validation images, and 1,200 testing images. This structured approach enabled a thorough and scalable evaluation of the model's segmentation performance on a novel and diverse multi-class histopathological dataset. Together, these three datasets allow for robust and diverse benchmarking of the segmentation system, ensuring its adaptability to various medical imaging challenges and underlying tissue heterogeneity.

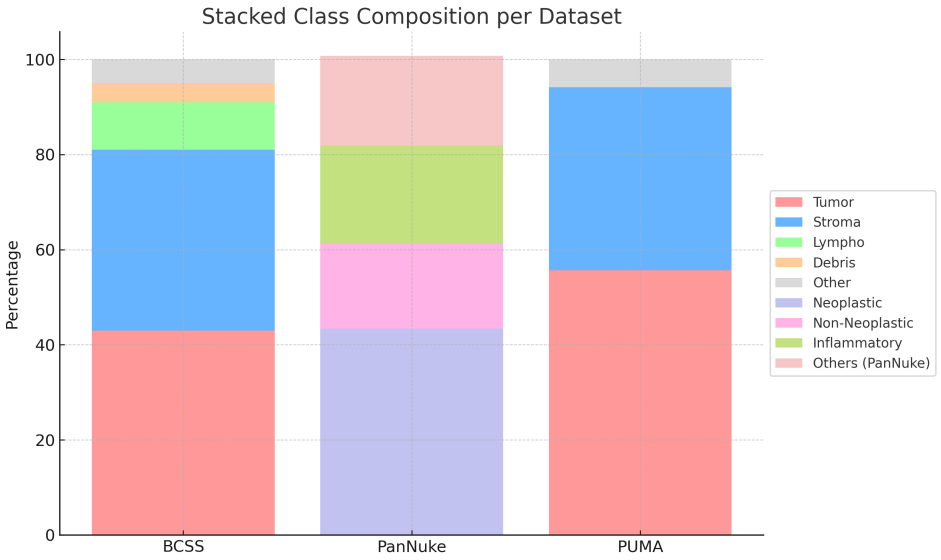


Figure 1. Datasets distribution.

Figure 1. provides a visual comparison of the class distributions across the three histopathology datasets used in this study: BCSS, PanNuke, and PUMA. Each dataset exhibits a distinct set of tissue classes with varying levels of representation. In the BCSS, the majority of pixels belonged to the tumor (43%) and stroma (38%) classes, followed by lymphocytic infiltration (10%), debris (4%), and other tissue (5%) classes. In contrast, PanNuke includes six heterogeneous nuclear categories, with neoplastic nuclei comprising the largest proportion (43.38%), followed by inflammatory (20.69%), non-neoplastic epithelial (17.81%), connective (11.42%), dead (5.41%), and non-nuclear (1.06%) nuclei. Similarly, the PUMA dataset is dominated by tumor (55.68%) and stroma (38.46%) regions, with other tissues accounting for only 5.86% of the total dataset. The dataset analysis revealed that the tumor and stroma classes dominate across all datasets, while minority classes (e.g., necrosis and lymphocytic infiltration) are less represented. In this study, we focus on the five most commonly analyzed tissue categories in BCSS—tumor, stroma, lymphocytic infiltration, necrosis/debris, and other tissue—as adopted in prior research [4, 25, 26, 27]. This class selection ensures comparability with existing studies while maintaining sufficient representation for stable model training. For a detailed explanation of the strategies adopted to mitigate class imbalance during model training, refer to Section 4.2.

3.2. Evaluation Metric

The evaluation metrics for the segmentation performance of the proposed detector were based on four widely recognized metrics. These key metrics were based on the counts of true positives (TPs), false positives (FPs), false

negatives (FNs), and true negatives (TNs) for all images in the dataset. True positive (TP) instances were identified as the intersections between the segmented cell membrane and the ground truth, whereas those that did not meet this criterion were categorized as false positives (FPs). False negatives (FNs) are determined by the missed portions of the ground truth, whereas true negatives (TNs) encompass image regions beyond the union of the segmentation and the ground truth.

3.2.1. Network accuracy metric It is a performance measure that quantifies the overall accuracy of the predictions made by the CNN. This is a comprehensive metric that considers both true-positive and true-negative predictions relative to the total number of instances in a dataset [41]. The network accuracy can be determined using

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

In the context of CNNs, accuracy is a fundamental metric that assesses the ability of a model to correctly classify instances in all classes. This approach provides a general overview of model effectiveness but may not be suitable for highly imbalanced datasets. In such cases, additional metrics, such as precision, recall, and F1 score, are often considered to provide a more nuanced evaluation of the model performance.

3.2.2. F1 score metric This is a measure of the model accuracy that balances both precision and recall [42]. It provides a single numerical value that considers both false positives and false negatives, thereby offering a comprehensive evaluation of model performance. The F1 score can be determined using

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2}$$

where

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

and

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

3.2.3. Mean Intersection over Union (MIoU) The mean intersection over union (MIoU) is a common evaluation metric used for segmentation tasks in computer vision [43]. It measures the overlap between the predicted segmentation and ground truth, providing an overall performance score for the model. The MIoU can be determined using

$$\text{MIoU} = \frac{1}{C} \sum_{i=1}^{C} \frac{P_i \cap G_i}{P_i \cup G_i} \tag{5}$$

where P_i is the predicted set of pixels for class i, G_i is the ground truth set of pixels for class i,9and C is the total number of classes. The intersection represents the common pixels between the prediction and ground truth, whereas the union represents the total number of pixels present in either the prediction or ground truth image. MIoU is the average IoU value for all classes.

3.2.4. Frequency weighted intersection over union (FwIoU) The frequency-weighted IoU (FwIoU) is a variant of the IoU metric that considers the frequency of each class in a dataset [43]. This metric is more important for classes that appear more frequently in datasets. The FwIoU can be determined using

$$\text{FwIoU} = \sum_{i=1}^{C} \frac{n_i}{N} \cdot \frac{P_i \cap G_i}{P_i \cup G_i} \tag{6}$$

where n_i is the number of pixels belonging to class i, and where N is the total number of pixels across all classes. This metric weighs the IoU of each class by its proportion in the dataset, thereby providing a balanced evaluation that considers the class imbalance.

3.3. The Proposed Model

We introduce CancerSeg-XA, a deep learning-based semantic segmentation model designed to address the challenges of segmenting complex cancer histopathological images. The model leverages multi-scale contextual information, spatial refinement, and attention mechanisms to achieve high segmentation accuracy across diverse tissue structures and cellular regions. Based on the DeepLabV3+ framework, CancerSeg-XA incorporates key architectural enhancements aimed at improving both generalization and precision of segmentation. Figure 2. provides an overview of the model architecture used in this study. To extract rich hierarchical features from the

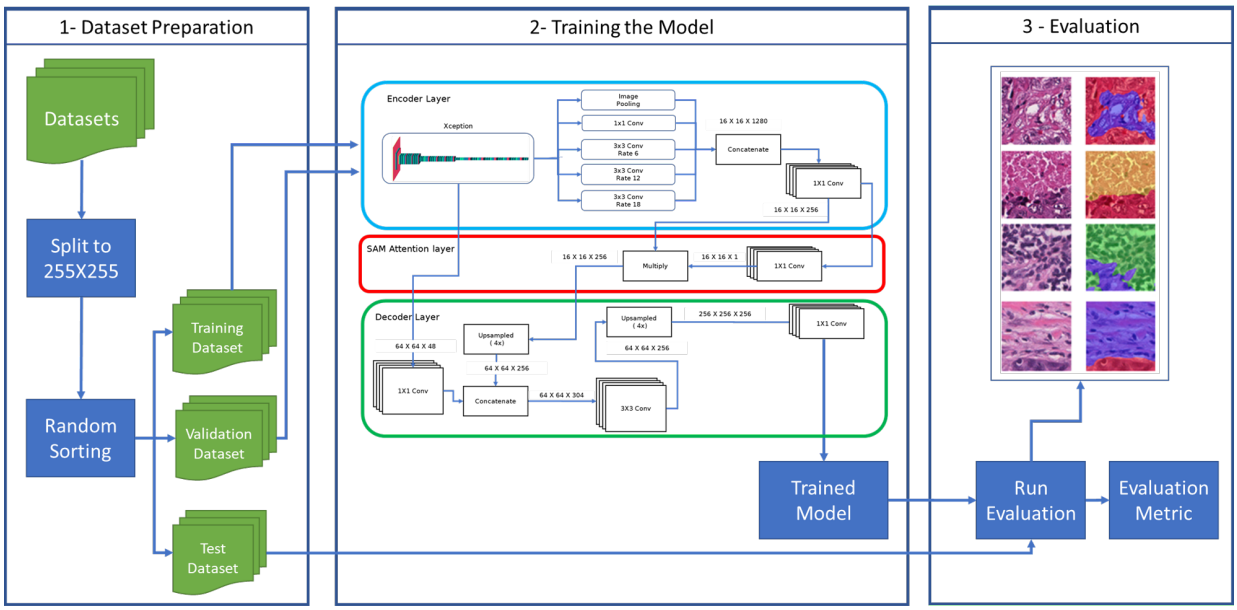


Figure 2. The proposed system modifications are based on the DeepLabV3+ model.

input images, the encoder component of the model replaces the original ResNet50 backbone of DeepLabV3+ with a more advanced Xception architecture. This substitution enhances the encoder's ability to capture localized, fine-grained features using depthwise separable convolutions. These characteristics are particularly beneficial for histopathological segmentation, where cellular and tissue variations are often subtle and difficult to detect. The improved encoder provides a more expressive feature representation, enabling better discrimination between complex tissue patterns and supporting downstream segmentation accuracy.

A spatial attention module (SAM) is inserted immediately after the Atrous Spatial Pyramid Pooling (ASPP) block and before the decoder input, i.e., the SAM takes the multi-scale encoder output as its input and produces an attended feature tensor that is forwarded to the decoder. Let $F \in \mathbb{R}^{H \times W \times C}$ denote the ASPP output feature tensor

(in our implementation $H = W = 16$ and $C = 1280$). The SAM computes a single-channel spatial attention map $A \in \mathbb{R}^{H\times W\times 1}$ by a 1×1 convolution followed by a sigmoid activation:

$$\mathrm{A} = \sigma\big(\mathrm{Conv}_{1\times 1}(F;\, W_a, b_a)\big) \tag{7}$$

where $\mathrm{Conv}_{1\times 1}(\cdot; W_a, b_a)$ denotes a 1×1 convolution parameterized by weights $W_a \in \mathbb{R}^{1\times 1\times C\times 1}$ and bias $b_a \in \mathbb{R}$, and $\sigma(\cdot)$ is the element-wise sigmoid function. The attention map is broadcast-multiplied across channels and applied element-wise to the original feature tensor to produce the attended features F':

$$\mathrm{F}' = F \odot A \tag{8}$$

where $\odot$ denotes element-wise multiplication with implicit channel-wise broadcasting of A to shape $H \times W \times C$. In our implementation the 1×1 conv produces an attention tensor of size $16 \times 16 \times 1$ which is multiplied with the ASPP output $16 \times 16 \times 1280$ to yield the attended tensor $16 \times 16 \times 1280$. This design adds only a small number of parameters (1,281 trainable parameters in our setup: C weights plus a bias), thereby preserving computational efficiency while improving spatial discrimination.

The decoder stage integrates features from multiple depths using skip connections to maintain spatial context while reconstructing the full-resolution segmentation map. These combined features are refined through convolutional operations and upsampling, ultimately producing a pixel-wise segmentation mask that assigns class labels to images. This architectural design ensures that the model captures detailed visual cues and delivers precise and interpretable segmentation results.

Through these enhancements, CancerSeg-XA demonstrates a strong capability of segmenting complex histopathological images, offering a practical and scalable solution for diverse cancer diagnostic tasks.

4. Results analysis

This section describes the training process and results.

4.1. Training setup

CancerSeg-XA was implemented in Python 3.7, using the Keras library as an interface for TensorFlow. The training was performed in the Kaggle environment, which provides a P100 GPU with 16 GB RAM and powerful cloud computing resources. The model used the Adam optimizer and sparse categorical cross-entropy loss function.

A reduced learning rate function was used to adjust the learning rate, with a reduction factor of 0.2, and a patience parameter was set to five epochs. In this study, the DeepLabV3+ model was trained for 25 epochs, which was consistent with the original configuration. Although we explored increasing the number of training epochs, the empirical results showed that the model began to overfit beyond epoch 25, without yielding further improvements. Therefore, we adhered to the original training schedule to maintain consistency and ensure a fair comparison with the proposed CancerSeg-XA model. The CNN models trained by tuning hyperparameters are shown in Table 1.

Table 1. Model training hyperparameter values

Hyperparameter	**DeepLabV3+**	**CancerSeg-XA**
Initial Learn Rate	1×10^{-2}	1×10^{-4}
End Learn Rate	1×10^{-4}	
No. of Epochs	25	38
Batch size	24	
Image size	256 x 256	
Trainable params	10,611,830	8,038,479
Nontrainable params	32,224	13,456

4.2. Loss Function and Class Imbalance Handling

As described in Section 3.1, the BCSS and PUMA datasets exhibit notable class imbalance, with tumor and stroma regions dominating the annotated samples. To address the issue of class imbalance inherent in histopathology datasets such as BCSS, PanNuke, and PUMA, a weighted cross-entropy loss function was employed during model training. This approach ensures that minority classes contribute proportionally to the overall loss, preventing the model from being biased toward dominant tissue types such as tumor and stroma. The class weights were computed based on the inverse frequency of each class in the training dataset.

The weighted cross-entropy loss $\mathcal{L}_{wce}$ is defined as:

$$\mathcal{L}_{wce} = -\sum_{i=1}^{C} w_i \, y_i \, \log(\hat{y}_i) \tag{9}$$

where C represents the total number of classes, y_i is the true label for class i, $\hat{y}_i$ denotes the predicted probability for that class, and w_i is the class weight determined by the inverse frequency of class occurrence:

$$w_i = \frac{1/f_i}{\sum_{j=1}^{C}(1/f_j)} \tag{10}$$

where f_i represents the frequency of class i in the training dataset. This weighting scheme assigns higher importance to underrepresented classes, improving the model's ability to learn fine-grained tissue structures across all histological categories. The application of this loss function ensured more balanced optimization and enhanced segmentation performance for minority classes such as necrosis, debris, and non-nuclear regions.

4.3. Training Progress

This section compares the training performances of DeepLabV3+ and CancerSeg-XA across the PanNuke, PUMA, and BCSS datasets, focusing on their final epochs: epoch 25 for DeepLabV3+ across all datasets, epoch 38 for CancerSeg-XA on PanNuke, and epoch 45 for CancerSeg-XA on the PUMA and BCSS datasets. The analysis examined the validation accuracy and loss, highlighting the influence of the learning rate (LR) schedule on the performance and stability of each model used.

For the BCSS dataset, as shown in Figure 3, DeepLabV3+ achieved a final validation accuracy of 80.00% at epoch 25, whereas CancerSeg-XA achieved 85.00% at epoch 45. The higher accuracy of CancerSeg-XA underscores the benefits of its prolonged training period. DeepLabV3+ required an early LR reduction to 2e-4 after initial instability, followed by further adjustments, which stabilized the training but limited the accuracy gains. CancerSeg-XA maintained its initial LR until epoch 27, with subsequent reductions, allowing sustained learning and finer optimization, resulting in a better performance in the final epoch.

For training validation loss, DeepLabV3+ recorded a final validation loss of 0.35 at epoch 25 on the BCSS, whereas CancerSeg-XA achieved 0.28 at epoch 45. The lower loss for CancerSeg-XA reflects its enhanced generalization, supported by extended training and an adaptive LR schedule. DeepLabV3+'s early LR adjustments prevented significant loss reduction beyond epoch 25, whereas CancerSeg-XA's delayed reductions (e.g., to 2e-4 at epoch 27) facilitated ongoing refinement, minimizing the loss more effectively over 45 epochs.

On the PanNuke dataset, as shown in Figure 4, DeepLabV3+ achieved a final validation accuracy of 85.57% at epoch 25, whereas CancerSeg-XA reached 87.00% at epoch 38. The higher accuracy of CancerSeg-XA demonstrates the advantage of its extended training, allowing further optimization beyond DeepLabV3+'s 25 epochs. Both models began with an initial learning rate (LR) of 1e-2. DeepLabV3+ reduced its LR to 1e-3 at epoch 9 and 1e-4 at epoch 20, stabilizing the training but limiting the further gains. CancerSeg-XA, with reductions at epochs 18 and 24, maintained a higher LR for a longer period, enabling more aggressive learning early on and finer adjustments later, contributing to its superior accuracy.

For the PanNuke validation loss, DeepLabV3+ recorded a final validation loss of 0.30 at epoch 25, whereas CancerSeg-XA achieved a lower loss of 0.25 at epoch 38. The extended training of CancerSeg-XA, paired with its LR schedule (reductions at epochs 18 and 24), allowed for better convergence and reduced overfitting compared

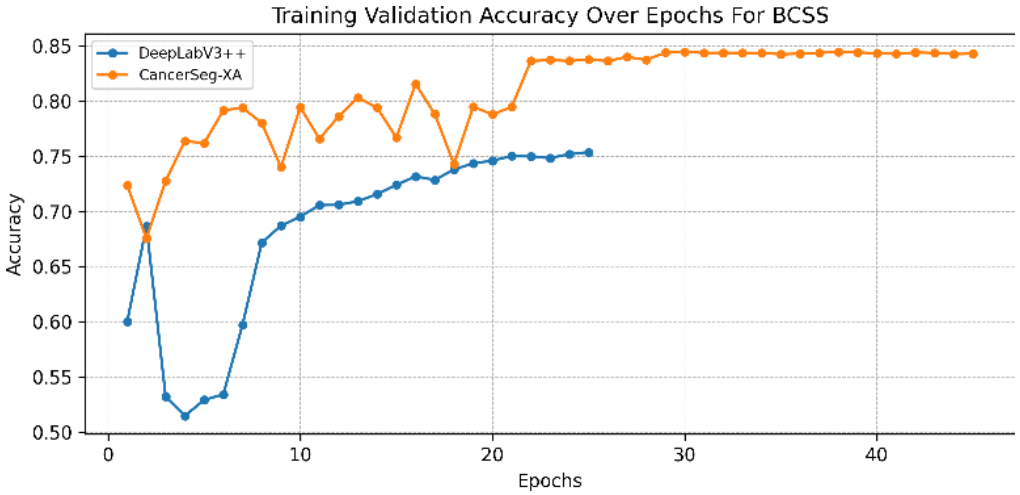


Figure 3. Validation accuracy for BCSS, comparing DeepLabV3+ at epoch 25 and CancerSeg-XA at epoch 45.

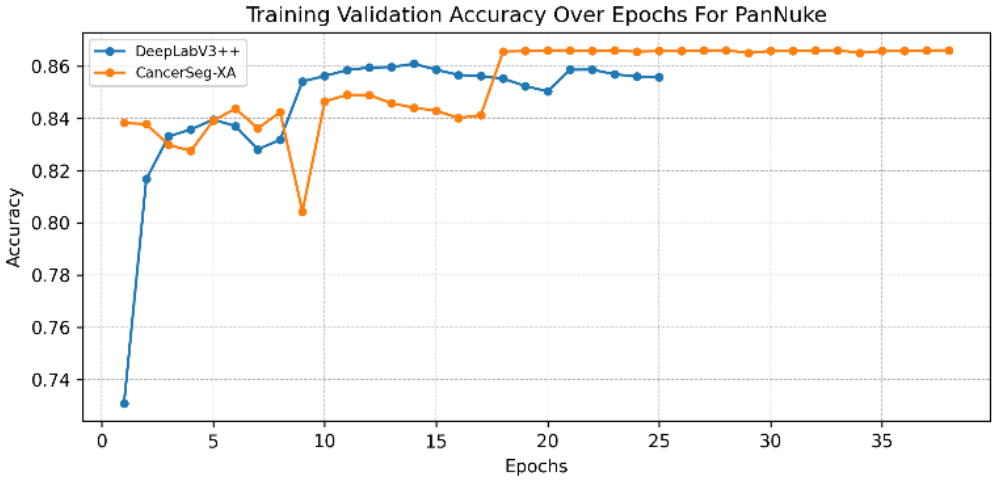


Figure 4. Validation accuracy for PanNuke, comparing DeepLabV3+ at epoch 25 and CancerSeg-XA at epoch 38.

with DeepLabV3+. DeepLabV3+'s earlier LR reductions stabilized its loss but prevented further improvement, whereas the gradual decay of CancerSeg-XA supported continued loss reduction, enhancing generalization.

On the PUMA dataset, as shown in Figure 5, DeepLabV3+ reached a final validation accuracy of 66.00% at epoch 25, whereas CancerSeg-XA achieved 72.26% at epoch 45. The significant improvement in CancerSeg-XA reflects its ability to leverage 20 additional epochs for optimization. DeepLabV3+ adjusted its LR at epochs 8 (to 1e-4), 14 (to 1e-5), and 21 (to 1e-6), stabilizing the training but causing an early plateau. CancerSeg-XA, with LR reductions at epochs 17, 28, and 37, benefited from a more gradual decay, maintained learning momentum, and achieved higher accuracy through its extended training and adaptive architecture.

This observation is further supported by the fact that DeepLabV3+ recorded a final validation loss of 0.45 at epoch 25 on PUMA, compared to CancerSeg-XA's 0.35 at epoch 45. The lower loss of CancerSeg-XA highlights its superior generalization, driven by extended training and a well-tuned LR schedule (reductions at epochs 17, 28, and 37). DeepLabV3+'s aggressive early LR reductions limited its ability to minimize loss further, whereas CancerSeg-XA's delayed and gradual adjustments enabled consistent improvement, effectively reducing loss over additional epochs.

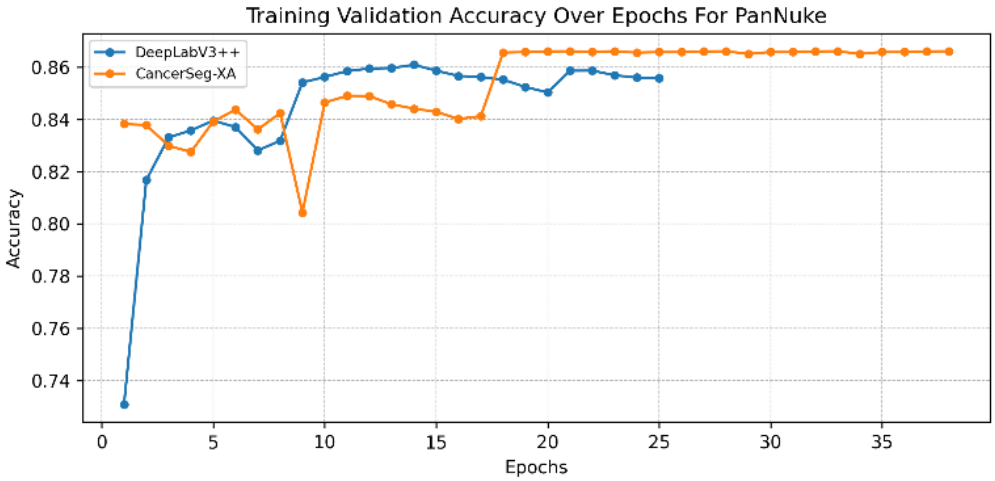


Figure 5. Validation accuracy for PUMA, comparing DeepLabV3+ at epoch 25 and CancerSeg-XA at epoch 45.

4.4. Statistical Validation of Model Performance

To ensure the robustness and statistical reliability of the results, each model was independently trained and evaluated three times with different random seeds, while maintaining identical hyperparameters and dataset splits. The mean and standard deviation of key performance metrics (Accuracy, Recall, F1-score, MIoU, and FwIoU) were computed to capture potential training variability. Additionally, paired t-tests were conducted to assess the statistical significance of CancerSeg-XA's improvements over the baseline DeepLabV3+, with a significance level set at $p < 0.05$.

Table 2 presents the statistical validation results on the BCSS dataset. CancerSeg-XA consistently outperformed DeepLabV3+ across all metrics with lower variance and statistically significant gains, confirming that the observed improvements are robust and not due to random variation.

Table 2. Statistical validation over three independent runs on the BCSS dataset

Metric	**DeepLabV3+ (Mean ± SD)**	**CancerSeg-XA (Mean ± SD)**	**p-value**
Accuracy	0.8670 ± 0.0093	0.9148 ± 0.0018	0.0158
Recall	0.8108 ± 0.0181	0.8819 ± 0.0092	0.0162
F1-score	0.8026 ± 0.0182	0.8796 ± 0.0080	0.0151
MIoU	0.7498 ± 0.0180	0.8282 ± 0.0105	0.0107
FwIoU	0.7442 ± 0.0169	0.8180 ± 0.0103	0.0104

To further verify the consistency of model performance, both DeepLabV3+ and CancerSeg-XA were independently trained and evaluated three times on the PanNuke dataset under identical configurations but with different random seeds. Table 3 summarizes the mean, standard deviation, and corresponding p-values for all major performance metrics. Similar to the BCSS dataset, CancerSeg-XA demonstrates superior performance and lower variance across all metrics. The p-values ($p < 0.05$) confirm that the observed improvements are statistically significant, underscoring the robustness of CancerSeg-XA in handling heterogeneous multi-organ histopathology images.

A similar statistical validation was performed for the PUMA dataset to ensure the consistency and reliability of the results. Each model was trained and evaluated across three independent runs. Table 4 presents the aggregated results. CancerSeg-XA consistently outperformed DeepLabV3+ in all key metrics, with accuracy, recall, and F1-score increasing by approximately 3.2%, and statistically significant gains observed across all measures ($p < 0.05$).

Table 3. Statistical validation over three independent runs on the PanNuke dataset

Metric	DeepLabV3+ (Mean ± SD)	CancerSeg-XA (Mean ± SD)	p-value
Accuracy	0.8556 ± 0.0110	0.8987 ± 0.0044	0.0199
Recall	0.8653 ± 0.0031	0.8891 ± 0.0030	0.0015
F1-score	0.8353 ± 0.0068	0.8760 ± 0.0046	0.0035
MIoU	0.7795 ± 0.0079	0.8261 ± 0.0044	0.0045

These findings reaffirm that CancerSeg-XA maintains stable and reliable segmentation performance even on datasets with pronounced variability and complex tissue structures.

Table 4. Statistical validation over three independent runs on the PUMA dataset

Metric	DeepLabV3+ (Mean ± SD)	CancerSeg-XA (Mean ± SD)	p-value
Accuracy	0.8787 ± 0.0089	0.9109 ± 0.0035	0.0242
Recall	0.8787 ± 0.0089	0.9109 ± 0.0035	0.0242
F1-score	0.8787 ± 0.0089	0.9109 ± 0.0035	0.0242
MIoU	0.5531 ± 0.0042	0.6147 ± 0.0103	0.0066
FwIoU	0.8230 ± 0.0114	0.8655 ± 0.0028	0.0285

4.5. Results

This section presents the detailed evaluation of CancerSeg-XA across the BCSS, PanNuke, and PUMA datasets. The comparative analysis highlights the performance improvements over the baseline DeepLabV3+ and other state-of-the-art methods.

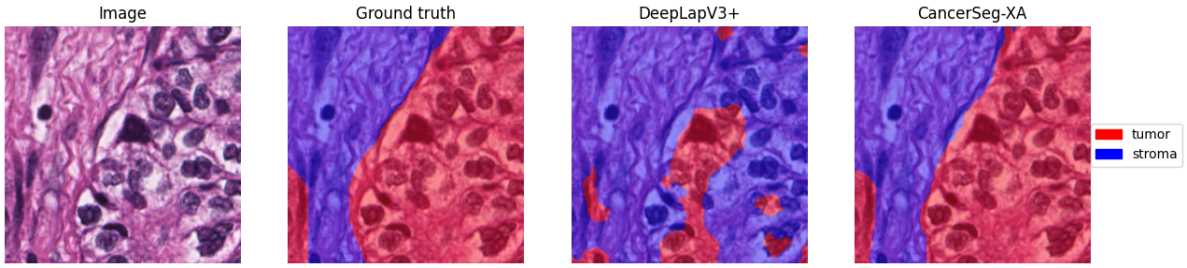
On the BCSS dataset, CancerSeg-XA was benchmarked against DeepLabV3+ as the baseline model. After a detailed performance evaluation, CancerSeg-XA achieved notable improvements over DeepLabV3+: accuracy by 4.79% (from 86.70% to 91.48%), recall by 7.11% (from 81.08% to 88.19%), F1-score by 7.71% (from 80.26% to 87.96%), MIoU by 7.84% (from 74.98% to 82.82%), and FwIoU by 7.38% (from 74.42% to 81.80%). These improvements confirm the strong segmentation capability of CancerSeg-XA, as shown in Table 5.

Table 5. Performance comparison of CancerSeg-XA and existing methods on the BCSS dataset

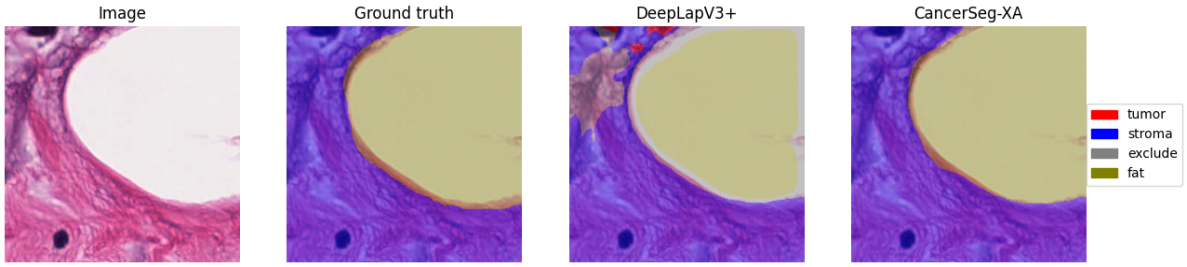
Model	Accuracy	recall	F1-score	MIoU	FwIoU
Han et al. [4]	0.84832	-	-	0.6892	0.7374
Srijay et al. [25]	0.77	-	-	-	-
Huang et al. [26]	0.8594	-	0.6989	-	-
Mauricio et al. [27]	0.81	-	-	-	-
DeepLabV3+	0.8670	0.8108	0.8026	0.7498	0.7442
CancerSeg-XA	0.9148	0.8819	0.8796	0.8282	0.8180

In addition to the baseline DeepLabV3+, CancerSeg-XA was evaluated against several state-of-the-art methods previously applied to the BCSS dataset, including the models proposed by Han et al. [4], Srijay et al. [25], Huang et al. [26], and Mauricio et al. [27]. CancerSeg-XA outperformed Han et al.'s model in terms of accuracy (91.48% vs. 84.83%) and FwIoU (81.80% vs. 73.74%). Compared with Huang et al., who achieved 85.94% accuracy and an F1-score of 69.89%, CancerSeg-XA delivered notable gains across all evaluated metrics, confirming its superior generalization capability. Furthermore, it significantly surpassed the models of Srijay et al. and Mauricio et al., who reported accuracies of 77% and 81%, respectively. These results affirm that the architectural enhancements in CancerSeg-XA—specifically the integration of the Xception backbone and spatial attention module—contribute to more robust and accurate segmentation performance on breast cancer histopathological images. These quantitative

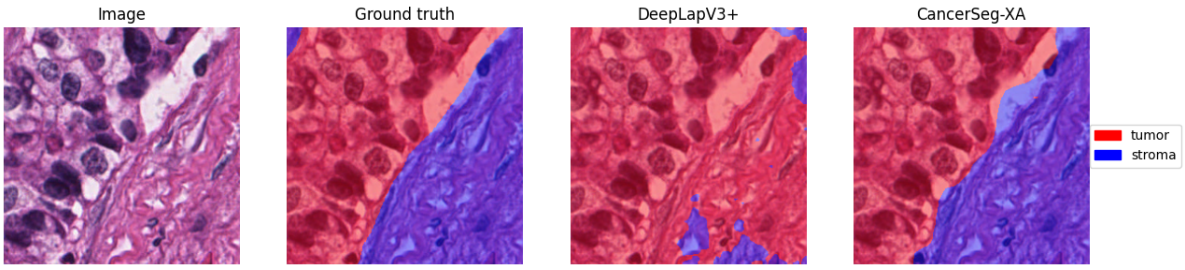
gains are also reflected qualitatively, as shown in Figure 6, where CancerSeg-XA demonstrates clearer boundary localization and substantially reduced misclassification compared with DeepLabV3+ across representative BCSS samples.



(a) **Example 1**: CancerSeg-XA delineates tumor-stroma boundaries with significantly higher precision and fewer fragmented regions than DeepLabV3+.



(b) **Example 2**: CancerSeg-XA better preserves the continuity of stromal regions and markedly reduces false-positive tumor predictions in ambiguous areas.



(c) **Example 3**: CancerSeg-XA accurately excludes fat and necrotic regions while maintaining crisp separation between tumor and surrounding stroma.

Figure 6. Qualitative segmentation results on the BCSS dataset. From left to right: original H&E image, ground truth, DeepLabV3+ prediction, and CancerSeg-XA prediction (red = tumor, blue = stroma, yellow = fat, gray = exclude). CancerSeg-XA consistently produces cleaner, more accurate tissue boundaries across diverse breast cancer histologies.

The PanNuke dataset was originally introduced for nuclei instance segmentation, but several recent studies have adapted it for semantic segmentation, enabling a fair comparison with our method. Based on this semantic-segmentation setting, CancerSeg-XA achieved clear improvements over all state-of-the-art (SOTA) models listed in Table 6. Compared with the baseline DeepLabV3+, CancerSeg-XA improved the F1-score from 0.8353 to 0.876 (+4.07%) and the MIoU from 0.7795 to 0.8261 (+4.66%), demonstrating stronger class-level consistency and spatial accuracy across heterogeneous tissue regions.

When benchmarked against other leading semantic-segmentation approaches, CancerSeg-XA also achieved the highest performance. Specifically, it outperformed SAM-ZeroShot by +2.17% in F1-score (0.876 vs. 0.8543) and surpassed SNOW by +2.88% in F1-score (0.876 vs. 0.8472) and +9.40% in MIoU (0.8261 vs. 0.7321). Relative to

SegFormer-b0 and SegFormer-b4, CancerSeg-XA achieved higher F1-scores by +5.0% and +3.08%, respectively. Finally, compared with the lightweight AttE-UNet, CancerSeg-XA achieved substantial gains of +11.3% in F1-score (0.876 vs. 0.763) and +17.2% in MIoU (0.8261 vs. 0.654). These improvements highlight the superior robustness and generalization capability of CancerSeg-XA for semantic segmentation on the challenging PanNuke dataset.

To further illustrate these quantitative improvements, Figure 7 presents qualitative examples from the PanNuke dataset, showing clearer nuclei delineation and reduced boundary ambiguity in CancerSeg-XA compared with DeepLabV3+ across multiple tissue regions.

Table 6. Performance comparison of CancerSeg-XA and existing methods on the PanNuke dataset

Model	**F1-score**	**MIoU**
SAM-ZeroShot [33]	0.8543	-
SNOW [35]	0.8472	0.7321
Segformer-b0 [36]	0.8260	-
Segformer-b4 [36]	0.8451	-
AttE-Unet [37]	0.763	0.654
DeepLabV3+	0.8353	0.7795
CancerSeg-XA	0.876	0.8261

For the PUMA dataset, after a detailed performance analysis, CancerSeg-XA demonstrated consistent and substantial improvements over the baseline DeepLabV3+. Specifically, it achieved higher accuracy by 3.22% (from 87.87% to 91.09%), recall by 3.22% (from 87.87% to 91.09%), F1-score by 3.22% (from 87.87% to 91.09%), MIoU by 6.16% (from 55.31% to 61.47%), and FwIoU by 4.25% (from 82.30% to 86.55%). These improvements confirm CancerSeg-XA's enhanced segmentation capability and robustness on the PUMA dataset. Furthermore, when compared to previously reported approaches, CancerSeg-XA significantly outperformed the PUMA Baseline model [40], which achieved an F1-score of 69.4%, and Torbati et al. [38], which reached 74.39%. This superior performance highlights CancerSeg-XA's generalization ability and its effectiveness across diverse histopathological domains, as detailed in Table 7. The qualitative examples in Figure 8 further illustrate these improvements, showing clearer tumor delineation and more reliable segmentation in complex morphological patterns.

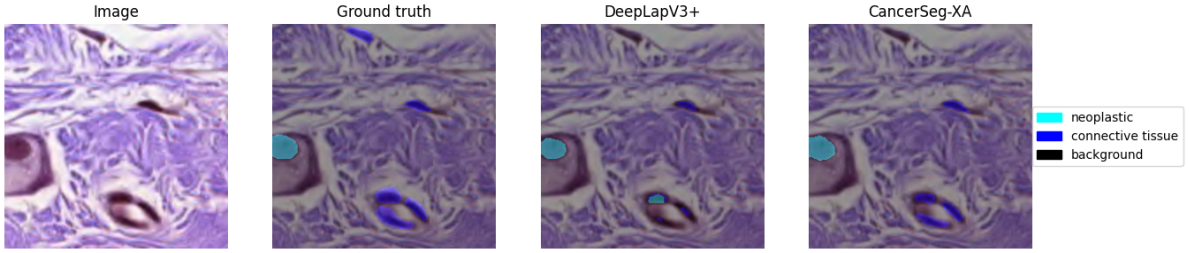
Table 7. Performance comparison of CancerSeg-XA and existing methods on the PUMA dataset

Model	**Accuracy**	**Recall**	**F1-score**	**MIou**	**FwIou**
PUMA Baseline [40]	-	-	0.694	-	-
Torbati et al. [38]	-	-	0.7439	-	-
DeepLabV3+	0.8787	0.8787	0.8787	0.559	0.8230
CancerSeg-XA	0.9109	0.9109	0.9109	0.6147	0.8655

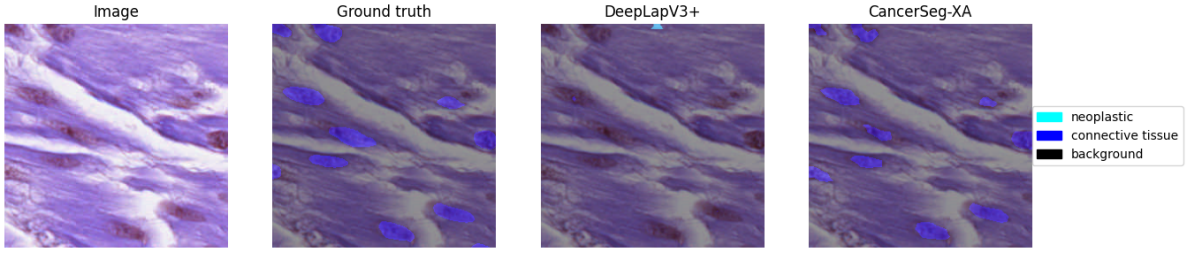
4.6. Impact of Architectural Modifications on Model Performance

The impact of the proposed architectural enhancements—namely the integration of a spatial attention layer and the replacement of the ResNet50 backbone with Xception—is summarized in Table 8. Replacing the ResNet50 encoder with Xception substantially reduced model complexity, decreasing the number of layers from 147 in DeepLabV3+ to 52 in CancerSeg-XA and reducing the model size by 24.2% (from 40.48 MB to 30.66 MB). The addition of the spatial attention module introduced only two extra layers and 1,281 trainable parameters (less than 0.01% of the total), yet contributed meaningfully to performance gains across all datasets.

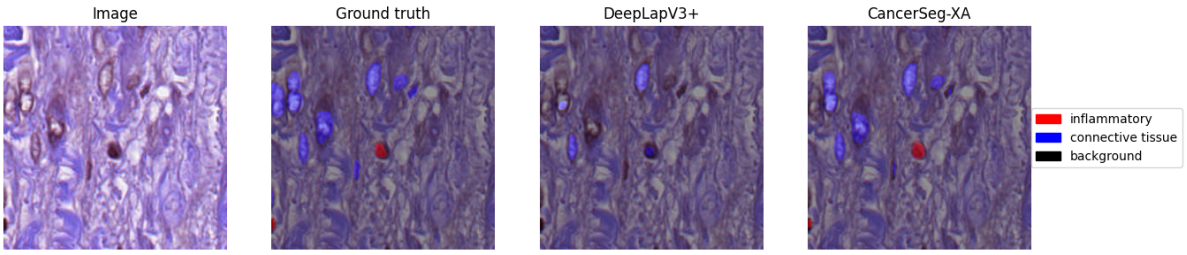
To evaluate computational efficiency in a clinically realistic setting, both models were benchmarked on a whole-slide-image (WSI) simulation consisting of batches of 20 patches—reflecting the common scenario in which WSIs are processed as moderate-sized patch sets during diagnostic workflows. Inference benchmarks were conducted on



(a) **Example 1**: CancerSeg-XA detects and delineates individual neoplastic nuclei with higher fidelity and far fewer false positives than DeepLabV3+.



(b) **Example 2**: In densely packed regions, CancerSeg-XA maintains better nuclear instance separation and correctly classifies inflammatory nuclei missed by DeepLabV3+.



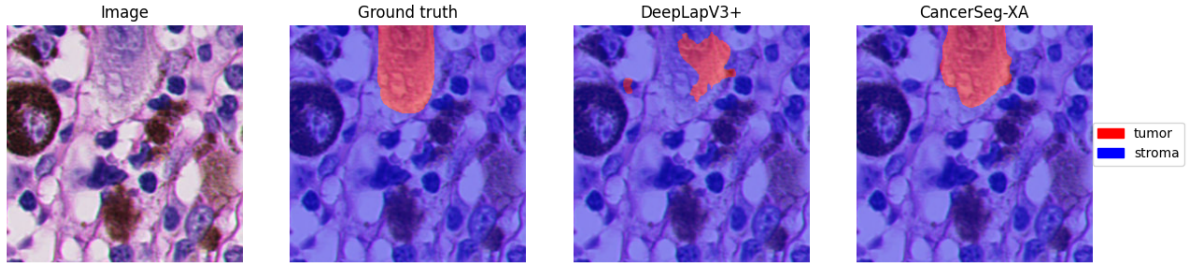
(c) **Example 3**: CancerSeg-XA robustly identifies small clusters of neoplastic nuclei in connective tissue while suppressing background noise more effectively.

Figure 7. Qualitative pan-cancer nuclei segmentation results on the PanNuke dataset. From left to right: original image, ground truth, DeepLabV3+ prediction, and CancerSeg-XA prediction (cyan = neoplastic, blue = connective tissue, red = inflammatory, black = background). CancerSeg-XA achieves superior nuclear detection and classification across 19 different tissue types.

a Kaggle Tesla P100-PCIE-16GB GPU, where memory usage is capped at approximately 15.9 GB due to hardware limits.

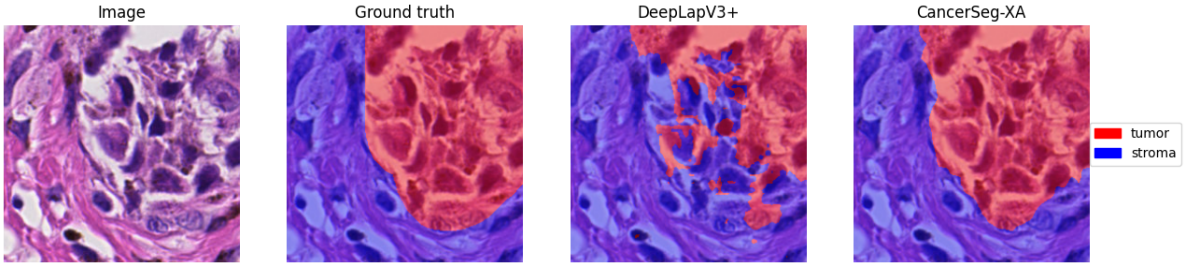
CancerSeg-XA demonstrated faster inference, achieving an average time of 0.4095 s per batch and 0.02048 s per image, compared to DeepLabV3+, which required 0.4454 s per batch and 0.02227 s per image. When extrapolated to a full WSI (approximately 20 patches), CancerSeg-XA reduced total processing time from 77.94 s to 71.67 s, representing a meaningful efficiency gain in the context of large-scale slide analysis. Standard deviation values further indicated lower temporal variability during inference, reflecting improved runtime stability.

These results collectively demonstrate that the architectural modifications not only enhance segmentation accuracy across all datasets but also improve computational efficiency and stability. This balance between accuracy, model compactness, and inference speed supports the practical feasibility of deploying CancerSeg-XA in real-world digital pathology pipelines, particularly in resource-constrained clinical environments.

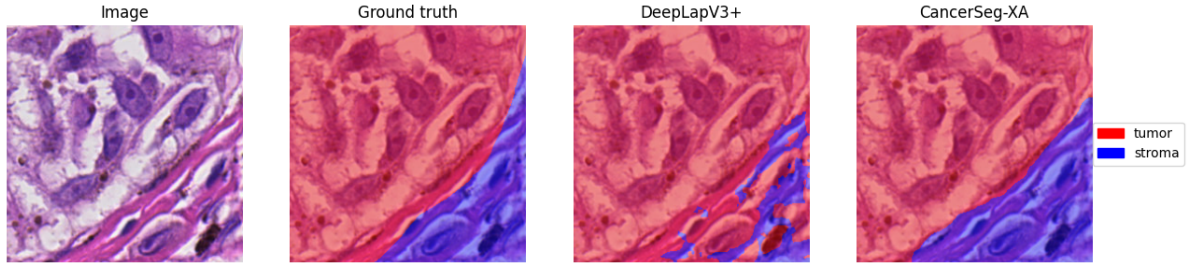


(a) **Example 1**: CancerSeg-XA captures fine tumor protrusions and thin stromal strands that DeepLabV3+ over-smooths or completely misses.

chassis



(b) **Example 2**: In highly irregular tumor morphologies, CancerSeg-XA preserves structural details and reduces leakage into surrounding stroma.



(c) **Example 3**: CancerSeg-XA exhibits with Xception+Attention clearly outperforms DeepLabV3+ in delineating complex, branching tumor regions with minimal fragmentation.

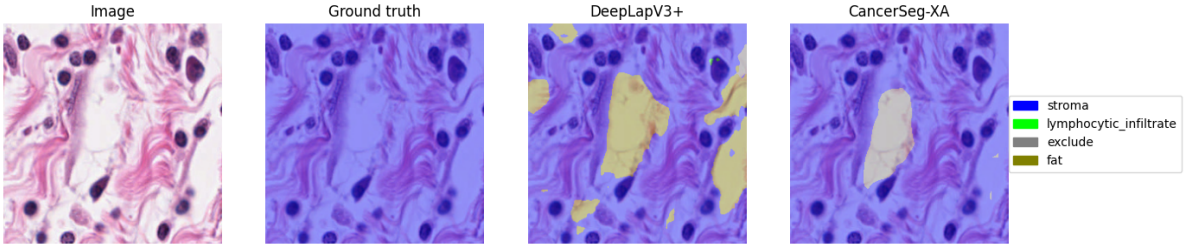
Figure 8. Qualitative segmentation results on the PUMA dataset. From left to right: original H&E image, ground truth, DeepLabV3+ prediction, and CancerSeg-XA prediction (red = tumor, blue = stroma). CancerSeg-XA exhibits superior boundary adherence and robustness on the challenging prostate cancer cases in PUMA.

Table 8. Computational and architectural comparison of DeepLabV3+ and CancerSeg-XA

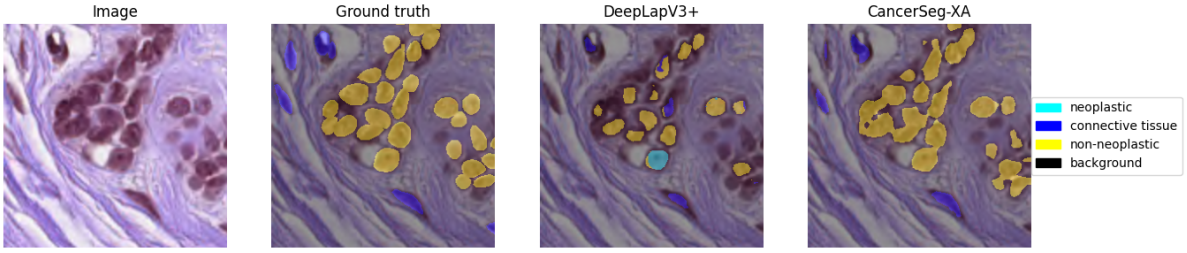
Model	Layers	Trainable Params	Size (MB)	Avg WSI Time
DeepLabV3+	147	10,611,830	40.48	0.4454 s
CancerSeg-XA	52	8,038,479	30.66	0.4095 s

4.7. Dataset-Driven Limitations and Histological Ambiguity

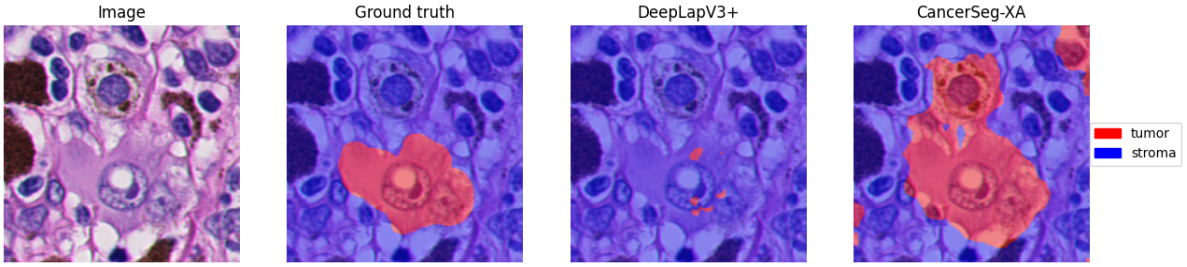
The qualitative examples in Figure 9 illustrate that the remaining discrepancies arise from inherent histopathology ambiguities rather than shortcomings of CancerSeg-XA. Across BCSS, PanNuke, and PUMA, many structures exhibit overlapping morphologies, subtle texture transitions, and visually indistinct class boundaries—particularly for rare categories such as isolated connective-tissue nuclei, small excluded regions, or transitional tumor–stroma



(a) BCSS example illustrating ambiguous neoplastic vs. non-neoplastic boundaries, overlapping nuclear morphologies, and subtle connective-tissue structures that contribute to segmentation variability.



(b) PanNuke example showing tumor–stroma continuity and nuanced nuclear textures that lead to plausible deviations in tumor boundary prediction.



(c) PUMA example highlighting subtle differences between stromal regions and pale extracellular or excluded spaces, complicating class separation for rare categories.

Figure 9. Qualitative error analysis across BCSS, PanNuke, and PUMA. The examples illustrate realistic histopathology challenges—ambiguous boundaries, overlapping morphological patterns, subtle textures, and rare classes—that contribute to discrepancies between CancerSeg-XA predictions and ground truth.

interfaces. In several regions, the ground-truth annotations rely on fine-grained, nucleus- or region-level delineations that exceed what is consistently visible in the raw tissue, making perfect agreement unattainable even for a strong model. Thus, the differences observed in Figure 9 reflect fundamental dataset and annotation constraints, not failures of CancerSeg-XA.

5. Conclusion

Cancer remains a critical global health challenge, necessitating the development of advanced computational tools to support early diagnosis and effective treatment planning. This study presents CancerSeg-XA, an enhanced deep learning model based on the DeepLabV3+ framework, which is designed for the semantic segmentation of cancer histopathological images. By integrating an attention layer and replacing the ResNet50 backbone with the Xception backbone, CancerSeg-XA achieved significant improvements in segmentation performance while

maintaining model efficiency. Evaluated on three diverse datasets (BCSS, PanNuke, and PUMA), CancerSeg-XA consistently outperformed the baseline DeepLabV3+. Specifically, it achieved accuracy improvements of 4.78%, 4.31%, and 3.22% on BCSS, PanNuke, and PUMA, respectively. Moreover, CancerSeg-XA demonstrated superior performance in terms of recall, F1-score, and frequency-weighted intersection over union (FwIoU), underscoring its robustness across various tissue types and imaging sources. The use of the Xception backbone not only enhanced the accuracy but also reduced the model complexity, resulting in a smaller size and fewer layers than the baseline. This efficiency is crucial for practical deployment in clinical practice. The consistent performance across multiple datasets highlights the ability of CancerSeg-XA to generalize effectively, making it a promising candidate for integration into digital pathology workflows. These findings demonstrate the potential of advanced architectures and attention mechanisms to transform histopathological image segmentation, paving the way for more accurate, efficient, and stable models that can support clinical decision-making and advance cancer care.

REFERENCES

[1] World Health Organization. Who - cancer. https://www.who.int/news-room/fact-sheets/detail/cancer. Accessed: 2021-09-26.

[2] American Institute for Cancer Research. Worldwide cancer data. https://www.wcrf.org/preventing-cancer/cancer-statistics/worldwide-cancer-data/, 2022. Accessed: 2025-07-25.

[3] Savannah R Duenweg, Samuel A Bobholz, Allison K Lowman, Margaret A Stebbins, Aleksandra Winiarz, Biprojit Nath, Fitzgerald Kyereme, Kenneth A Iczkowski, and Peter S LaViolette. Whole slide imaging (wsi) scanner differences influence optical and computed properties of digitized prostate cancer histology. *Journal of pathology informatics*, 14:100321, 2023.

[4] Chu Han, Jiatai Lin, Jinhai Mai, Yi Wang, Qingling Zhang, Bingchao Zhao, Xin Chen, Xipeng Pan, Zhenwei Shi, Zeyan Xu, et al. Multi-layer pseudo-supervision for histopathology tissue semantic segmentation using patch-level classification labels. *Medical Image Analysis*, 80:102487, 2022.

[5] Kunio Doi. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*, 31(4-5):198–211, 2007.

[6] Asifullah Khan, Anabia Sohail, Umme Zahoora, and Aqsa Saeed Qureshi. A survey of the recent architectures of deep convolutional neural networks. *Artificial intelligence review*, 53(8):5455–5516, 2020.

[7] R Krithiga and P Geetha. Breast cancer detection, segmentation and classification on histopathology images analysis: a systematic review. *Archives of Computational Methods in Engineering*, 28(4), 2021.

[8] Alaa Mohamed Youssef, El Behaidy, Wessam Hassan, and Aliaa Abdel-Haleim Abdel-Razik Youssif. Cancerseg-xa: Enhanced breast cancer histo-pathology segmentation using xception backbone with attention mechanisms. *Journal of Communications Software and Systems*, 21(1):79–89, 2025.

[9] Jonathan De Matos, Alceu de Souza Britto Jr, Luiz ES Oliveira, and Alessandro L Koerich. Histopathologic image processing: A review. *arXiv preprint arXiv:1904.07900*, 2019.

[10] Asha Das, Madhu S Nair, and S David Peter. Computer-aided histopathological image analysis techniques for automated nuclear atypia scoring of breast cancer: a review. *Journal of digital imaging*, 33(5):1091–1121, 2020.

[11] Ivana Prvulović, Ika Kardum-Skelin, Dunja Susterčić, Jasminka Jakić-Razumović, and Spomenka Manojlović. Morphometry of tumor cells in different grades and types of breast cancer. *Collegium antropologicum*, 34(1):99–103, 2010.

[12] Dario Sitnik and Ivica Kopriva. Lefm-nets: Learnable explicit feature map deep networks for segmentation of histopathological images of frozen sections, 2022.

[13] MM Fraz, Muhammad Shaban, Simon Graham, Syed Ali Khurram, and Nasir M Rajpoot. Uncertainty driven pooling network for microvessel segmentation in routine histology images. In *International Workshop on Ophthalmic Medical Image Analysis*, pages 156–164. Springer, 2018.

[14] Yuchen Sun and Caicheng Shi. Liver tumor segmentation and subsequent risk prediction based on deeplabv3+. 612(2):022051, 2019.

[15] Sridevi K Subramanya, Rui Li, Ying Wang, Hiroshi Miyamoto, and Feng Cui. Deep learning for histopathological segmentation of smooth muscle in the urinary bladder. *BMC Medical Informatics and Decision Making*, 23(1):122, 2023.

[16] Hao Li, Jia Zhang, Jia Wang, Zhongke Feng, Boyi Liang, Nina Xiong, Junping Zhang, Xiaoting Sun, Yibing Li, and Shuqi Lin. Extracting citrus in southern china (guangxi region) based on the improved deeplabv3+ network. *Remote Sensing*, 15(23):5614, 2023.

[17] Javaria Amin, Muhammad Sharif, Steven Lawrence Fernandes, Shui-Hua Wang, Tanzila Saba, and Amjad Rehman Khan. Breast microscopic cancer segmentation and classification using unique 4-qubit-quantum model. *Microscopy Research and Technique*, 85(5):1926–1936, 2022.

[18] Sridevi Kayyur Subramanya. Deep learning models to characterize smooth muscle fibers in hematoxylin and eosin stained histopathological images of the urinary bladder, 2021.

[19] Jelena Musulin, Daniel Štifanić, Ana Zulijani, Tomislav Ćabov, Andrea Dekanić, and Zlatan Car. An enhanced histopathology analysis: An ai-based system for multiclass grading of oral squamous cell carcinoma and segmenting of epithelial and stromal tissue. *Cancers*, 13(8):1784, 2021.

[20] Yiheng Ju, Longbo Zheng, Peng Zhao, Fangjie Xin, Fengjiao Wang, Yuan Gao, Xianxiang Zhang, Dongsheng Wang, and Yun Lu. Artificial intelligence recognition of pathological t stage and tumor invasion in rectal cancer based on large panoramic pathological sections. *Intelligent Medicine*, 2(03):141–151, 2022.

[21] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[22] Patrick Nicholas Hadinata, Djoni Simanta, Liyanto Eddy, and Kohei Nagai. Crack detection on concrete surfaces using deep encoder-decoder convolutional neural network: a comparison study between u-net and deeplabv3+. 7(3):323–334, 2021.

[23] Yingci Liu, Elizabeth Bilodeau, Brian Pollack, and Kayhan Batmanghelich. Automated detection of premalignant oral lesions on whole slide images using convolutional neural networks. *Oral Oncology*, 134:106109, 2022.

[24] Mohamed Amgad, Habiba Elfandy, Hagar Hussein, Lamees A Atteya, Mai AT Elsebaie, Lamia S Abo Elnasr, Rokia A Sakr, Hazem SE Salem, Ahmed F Ismail, Anas M Saad, et al. Structured crowdsourcing enables convolutional segmentation of histology images. *Bioinformatics*, 35(18):3461–3467, 2019.

[25] Srijay Deshpande and Durga Parkhi. Spadesegresnet: Harnessing spatially-adaptive normalization for breast cancer semantic segmentation. In *Annual Conference on Medical Image Understanding and Analysis*, pages 344–356. Springer, 2024.

[26] Chenxi Huang, Shijia Liao, and Yonghong Peng. A two-stage weakly supervised semantic segmentation model based on pathological tissue relationships, 2023.

[27] Mauricio Alberto Ortega-Ruíz, Cefa Karabağ, Edgar Roman-Rangel, and Constantino Carlos Reyes-Aldasoro. Drd-unet, a unet-like architecture for multi-class breast cancer semantic segmentation. *IEEE Access*, 12:40412–40424, 2024.

[28] Yu-Han Im, Seo-Hyeong Park, and Sang-Chul Lee. Hda-net: H&e and rgb dual attention network for nuclei instance segmentation. *IEEE Access*, 12:56622–56632, 2024.

[29] Oscar Pina, Eduard Dorca, and Verónica Vilaplana. Cell nuclei detection and classification in whole slide images with transformers, 2025.

[30] Roua Jaafar, Hedi Yazid, Wissem Farhat, and Najoua Essoukri Ben Amara. Sbc-unet3+: Classification of nuclei in histology imaging based on multi branch unet3+ segmentation model. volume 601, page 609, 2025.

[31] Ibtihaj Ahmad, Zain Ul Islam, Saleem Riaz, and Fuzhong Xue. Mans-net: multiple attention-based nuclei segmentation in multi organ digital cancer histopathology images. *IEEE Access*, 12:173530–173539, 2024.

[32] Fabian Hörst, Moritz Rempe, Lukas Heine, Constantin Seibold, Julius Keyl, Giulia Baldini, Selma Ugurel, Jens Siveke, Barbara Grünwald, Jan Egger, et al. Cellvit: Vision transformers for precise cell segmentation and classification. *Medical Image Analysis*, 94:103143, 2024.

[33] Abu Bakor Hayat Arnob, Xiangxue Wang, Yiping Jiao, Xiao Gan, Wenlong Ming, and Jun Xu. Pathological primitive segmentation based on visual foundation model with zero-shot mask generation. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1–4. IEEE, 2024.

[34] Sai Chandana Koganti, Siri Yellu, Jihoon Yun, and Sanghoon Lee. Task-ready pannuke and nucls datasets: Reorganization, synthetic data generation, and experimental evaluation. *IEEE Access*, 2025.

[35] Kexin Ding, Mu Zhou, He Wang, Olivier Gevaert, Dimitris Metaxas, and Shaoting Zhang. A large-scale synthetic pathological dataset for deep learning-enabled segmentation of breast cancer. *Scientific Data*, 10(1):231, 2023.

[36] Marwan Khaled, Mostafa A Hammouda, Hesham Ali, Mustafa Elattar, and Sahar Selim. Efficient semantic segmentation of nuclei in histopathology images using segformer. In *Annual Conference on Medical Image Understanding and Analysis*, pages 81–95. Springer, 2023.

[37] Shuang Xia, Qian Sun, Yiheng Zhou, Zhaoyuxuan Wang, Chaoxing You, Kainan Ma, and Ming Liu. A lightweight neural network for cell segmentation based on attention enhancement. *Information*, 16(4):295, 2025.

[38] Nima Torbati, Anastasia Meshcheryakova, Ramona Woitek, Sepideh Hatamikia, Diana Mechtcheriakova, and Amirreza Mahbod. A multi-stage auto-context deep learning framework for tissue and nuclei segmentation and classification in h&e-stained histological images of advanced melanoma, 2025.

[39] Jevgenij Gamper, Navid Alemi Koohbanani, Ksenija Benet, Ali Khuram, and Nasir Rajpoot. Pannuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification. In *European congress on digital pathology*, pages 11–19. Springer, 2019.

[40] Mark Schuiveling, Hong Liu, Daniel Eek, Gerben E Breimer, Karijn PM Suijkerbuijk, Willeke AM Blokx, and Mitko Veta. A novel dataset for nuclei and tissue segmentation in melanoma with baseline nuclei segmentation and tissue segmentation benchmarks. *GigaScience*, 14:giaf011, 2025.

[41] Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

[42] Cornelis Joost Van Rijsbergen. *Information retrieval*. 1979.

[43] Alberto Garcia-Garcia, S Orts-Escolano, S Oprea, V Villena-Martinez, and J Garcia-Rodriguez. A review on deep learning techniques applied to semantic segmentation. arxiv 2017, 2020.