Journal of Theoretical and Applied Information Technology
15st September 2026. Vol.104. No.17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

TOWARDS PRECISION ONCOLOGY: ENHANCED MULTISCALE DEEP FEATURE LEARNING FOR CERVICAL CELL SEGMENTATION

JHEELAM MONDAL[1], RAJDEEP CHATTERJEE[1] , MAHENDRA KUMAR GOURISARIA[1,*]
[1]School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar-751027, Odisha, India
E-mail: [1]mkgourisaria2010@gmail.com

ABSTRACT

Over 300,000 people die from cervical cancer, which is the fourth most frequent malignancy in women worldwide. The early detection of cervical cancer corelates with significantly improved survival rates and the disease is largely preventable. It is indeed, a rare disease in affluent countries with robust immunization and screening initiatives. Nevertheless, the disease disproportionately impacts women in low- and middle-income nations, who often endure severe and untreatable diseases due to resource scarcity. Our objective is to create a reliable deep learning framework capable of distinguishing between normal and malignant morphologies by accurately delineating the cytoplasm and nucleus of the cells. We employ an innovative U-Net model named YOLOv5++ on a publicly accessible annotated dataset of cervical cytology images. Preliminary findings show the accuracy, IoU and Dice score values of 98.33%, 69.68% and 82.13% respectively which offers enhanced accuracy and efficiency compared to various other advanced U-Net methodologies. Thereby, our study demonstrates the enormous potential of deep learning-enabled segmentation as an effective technique to improve and standardize the cervical cancer screening process, reducing human error and enabling early clinical intervention.

Keywords: *Medical Image Analysis, Cervical Cancer, Pap Smear Images, U-Net Model, Image Segmentation*

1. INTRODUCTION

Cervical cancer is the fourth most prevalent type of cancer in women. Cervical cancer has a big effect on women's health and lives all around the world. The World Health Organization reported that over 6,60,000 women were diagnosed with cervical cancer in 2022, and 3,50,000 women globally succumbed to the disease [1]. Cervical cancer deaths happened mostly in countries with low and moderate incomes. The Pap smear test makes it possible to find and screen the cervical cancer early. This test is typically used to determine if a cell is healthy or malignant [2]. Expert doctors perform the smear analysis, which is tedious, time-consuming and error-prone process. In these developing and underdeveloped nations, the low output of cervical cancer screening procedures is caused by a lack of resources and highly qualified healthcare workers. The smear analysis takes a lot of time and work and is an error-prone procedure carried out by skilled medical professionals. Lack of resources and highly qualified healthcare doctors is the reason for the low output of cervical cancer screening methods in these emerging and underdeveloped countries. The presence of overlapping nuclei, shallow cells, weak contrast, false borders, poor staining and cytoplasm are some of the characteristics that still make this task difficult. Deep learning, which has lately become popular in medical domain, is the inspiration for the researchers [3]. A new multi-disciplinary area of study and advancement is health engineering, which is driven by the emergence of Big Data [4] and Machine Learning/Deep Learning [5-6]. Consequently, several deep-learning based methodologies have been developed for the diagnosing of cervical cancer.

Segmentation is the most often addressed topic in research using deep learning in medical imaging. It is generally regarded as an essential stage prior to disease diagnosis. Convolutional Neural Networks (CNNs) are one of the segmentation techniques that may be used to enormous datasets. In the study by Long et al. [7], fully convolutional network (FCN) based models are frequently employed for segmentation jobs because they are made for pixel-by-pixel classification. To produce segmentation maps, they use convolutional layers in place of the fully linked layers used in conventional CNNs. As

per study by Ronneberger et al. [8], models based on U-Net is a popular design developed for medical image segmentation. It is an encoder-decoder architecture with skip connections that improves segmentation accuracy while maintaining spatial information. Pan et al. [9] in their research, has enhanced U-Net by using atrous depthwise separable convolution (AS-UNet) for cell (nuclei) segmentation. The encoder, decoder and atrous convolution modules make up the AS-UNet. As the decoder module progressively retrieves the spatial data, the encoder module gathers the cell image's high-level semantic information layer by layer. In order to give the model a good perception capacity for both small and large cells, it extracts and merges multi-scale characteristics. Also, log-dice loss and focal loss are merged during the training phase and the network is optimized using the Adam optimizer approach. Cervical cell segmentation is a common application in diagnosing cervical cancer, which have gained prominence recently. Jung et al. [10] in their study demonstrated that deep convolutional neural networks (DCNN) perform exceptionally well when used to automatically extract features from unprocessed image data. Motivated by these successes, they have suggested a DCNN-based nuclei segmentation technique. A deep convolutional Gaussian mixture color normalization model that can cluster pixels has been used, while taking nuclear structures into account, to normalize the color of histopathology images. Mask R-CNN is employed to achieve cutting-edge object segmentation performance in the field of computer vision. For radiologists or clinical specialists, the main concerns are the segmentation, selection and extraction of the contaminated tumor from Magnetic Resonance Imaging (MRI) images. However, it takes a lot of time and effort, and its correctness is solely dependent on their experience. Sikder et al. [11] in their research proposes a novel approach that uses supervised learning, convolutional neural networks (CNNs) and morphological procedures to separate, categorize and identify different types of cancer cell from both MRI and RGB images. This approach uses semantic segmentation to divide the cancer cells and CNN to categorize different cancer types. In the review work by Wubineh et al. [12], the primary subjects are cell image segmentation and classification approaches for recognizing pap smear images for cervical cancer detection. The most popular methods for segmenting pap smear images among the chosen research are U-Net and Mask RCNN, while SVM and CNN are used to categorize their levels according to the derived features. Early detection strategy for mouth cancer has been explored using a unique deep learning framework that integrates morphological reconstruction with a UNet architecture by Devi et al. [13]. This framework attempts to obtain more precise segmentation of malignant regions by combining these two potent methodologies which will result in a more trustworthy classification of oral cancer type. Similar methodology is attained by Li et al. [14], where U-Net++ network is used as the segmentation backbone to give binary outputs which in turn is added to a VGG style architecture named RepVGG classifier to classify the images into labels. Following are the key points of our research contribution.

(1) Custom encoder architecture - For reliable feature extraction, an encoder is created from scratch utilizing the YOLOv5 backbone.
(2) Integration of C2f blocks of YOLOv8 – Conventional CSP blocks are replaced with effective C2f blocks of YOLOv8 thereby enhancing feature learning capabilities.
(3) Multi-Scale Deformable Attention Mechanism – We have applied adaptive attention to various feature scales and is effective in capturing complex spatial relationships present in the data.
(4) Encoder-Decoder architecture with skip connections – retrieves spatial data that is lost during downsampling and also preserves the encoder's fine-grained features by using skip connections.
(5) Optimized Training Strategy – Adam optimizer is used to achieve steady convergence and flexible learning rate adjustment. Autocast and mixed precision training is used to increase efficiency.

The subsections of this study are arranged as follows: Section 2 presents the literature review, Section 3 provides materials and methods used in this study, the results and discussion are presented in Section 4 along with the aspects of the proposed study that link to the other related studies. The conclusion and future work are covered in Section 5.

2. LITERATURE REVIEW

Geetha et al. [15] in their work used active contour approach to get an improved segmented image. ACM (snakes) is used in a variety of methods including locating object constraints and

recognizing edge segments that subsequently associate with one another. An optimization method called the analytic hierarchy process has been used to determine the weight factor for the energy minimization. The segmented regions are then used to extract the effective characteristics. Then, to improve classification accuracy, Teaching Learning based Optimization (ENN-TLBO) hybrid machine learning model and the Elman Neural Network has been used. This study achieved a classification accuracy of 89.60%. Doctors will be able to diagnose cervical cancer more quickly and accurately, if images related to colposcopy can be immediately used to autonomously segment the precancerous lesion region. Yuan et al. [16] has created three deep learning-based models using the private dataset. One saline image, one acetic image, one iodine image and the associated clinical data such as age, cytology and HPV test results, transformation zone type and pathologic diagnosis are gathered for each case. The dataset is divided into three parts; the training set, the test set and the validation set, using an 8:1:1 ratio. Model performance is assessed using the validation set. With an average accuracy of 95.59%, the segmentation model's DICE value for identifying lesions in acetic images is 61.64%. Additionally, the acetic detection model identified 84.67% of high-grade lesions. The diagnostic system outperformed colposcopists in standard colposcopy images, but did not perform well in high-definition images.

The research study by Liu et al. [17] suggests an enhanced technique for segmenting cervical squamous intraepithelial lesions using U-Net network. To enhance the U-Net encoder structure's feature extraction capabilities, a sub-coding module is first integrated into each layer. Second, to address the issue that the gradient vanishes as the network gets deeper, dense connections are added to the lowest level convolution operation. In this study, a new dataset is created from relevant hospital data and the model's precision, dice coefficient, IoU and recall indexes are superior to those of U-Net, FCN and SegNet. Results show that the highest dice value and IoU value achieved is 77.31% and 62.75% respectively. Zhao et al. [18] in their study provides a novel nuclei segmentation technique based on selective-edge-enhancement (SEENS). The suggested approach automatically avoids repeated segmentation and removes non-nuclei regions by combining mathematical operations with selective search to break up full-slide cervical images into smaller areas of interest (ROI). Additionally, edge information is extracted as a weight to selectively enhance the nucleus edge using an edge enhancement technique based on mathematical morphology and the canny operator. Consequently, the Chan-Vese model then more accurately segments the improved ROI. This method is assessed using 18 whole slide pictures of 395 cell nuclei. Results from experiments show that SEENS is more accurate at segmenting cervical nuclei.

Fan et al. [19] in their work adjusted the training loss and altered the U-Net structure in order to increase the accuracy of small target detection. Target boundaries are effectively handled by the changed structure when performing operations like bilinear upsampling. So, the modified U-Net includes shifting the skip connection locations, substituting transpose convolution with bilinear upsampling and also adding a novel loss function to address the challenge of small object recognition. Finally, the suggested approach is assessed using the dataset and contrasted with a number of deep learning-based techniques. The suggested strategy provides some advantages, and achieves a dice value of 87.6% and IoU value of 81.4%. The study by Chatterjee et al. [20] helps medical image processing researchers choose the U-Net that produce reliable results. Recent advances in the field of U-Nets include condensed U-Net, DU-Net and DRU-Net. DRU-Net which combines ResNet and DenseNet with a skip connection, has the best accuracy rate. It has attained 90.66% accuracy, 82.45% recall and 85.45% precision using 1850 images out of 4049 images of the SIPAKMED dataset. DRU-Net can be used in the future to segment overlapped cells in medical pictures.

Zhao et al. [21] have introduced a new cervical cancer segmentation dataset which contains three subsets: the domain segmentation dataset (DomainSeg) with data from various domains processing overlapping single nuclei images, the cluster segmentation dataset (ClusterSeg) with cluster nuclei and the patch segmentation dataset (PatchSeg) with nucleus images gathered under difficult conditions. The four metrics used to assess the performance of the nucleus segmentation network are the Hausdroff Distance (HD), Panoptic Quality (PQ), Aggregated Jaccard Index (AJI) and Aggregated Dice (Adice). A number of traditional as well as instance segmentation methods are compared. Also, on ISBI dataset, the popular techniques are evaluated. It is observed as per the experiments; the newly added public dataset can more accurately assess nucleus segmentation techniques. Ji et al. [22] in their work did automatic cervical cell segmentation using the Cx22 dataset.

Baseline model designs included U-Net, U-Net++, DeepLabV3, DeepLabV3Plus, Transunet and Segformer. The first four architectures used two separate encoders that could choose among ResNet34, ResNet50 and DenseNet121. Two methods are used to train the models; either from scratch or with encoders initialized from ImageNet pretrained models, followed by fine-tuning of every layer. When transfer learning is used, U-Net and U-Net++ with ResNet34 and DenseNet121 encoders usually do better than other methods. The dice similarity coefficient, sensitivity and specificity of baseline models for cytoplasm and nucleus segmentation are 94.80%, 95.40%, 98.23% and 75%, 71.30%, 99.88% respectively. All measures performed moderately better than the best baseline models, with the exception of cytoplasm segmentation specificity. Chowdhury et al. [23] has suggested a segmentation model based on residual-squeeze-and-excitation module, a model for feature extraction that uses fusion and a multi-layer perceptron classification model. Using the pretrained VGG19, VGG-F and CaffeNet models, the fusion-based feature extraction model gets three sets of deep features from these segmented nuclei. Two personally developed descriptors, bag-of-features and linear-binary-patterns are extracted for every image. The segmentation network has improved precision and ZSI by 2.04% and 2.00% on the Herlev dataset and by 0.68% and 2.59% on the ISBI2014 dataset.

Rudiansyah et al. [24] in their study produced an RMSE near 20% and IoU value of 78% during the learning period. The study's average precision, recall and F1 score are 89%,89% and 88.67% respectively, with an accuracy value of 89%. Their work demonstrates how effective the CNN U-Net architecture is at segmenting cervical cells for the nucleus and cytoplasm when combined with the enhanced image quality and data augmentation. Desiani et al. [25] in their study combines segmentation and augmentation. Augmentation combines pixel-wise transformation with gamma correction and geometric transformation with flipping and rotation. The suggested method produces a pap smear image with just two components; the background and the nuclei in the foreground. Using the augmented data, the standard U-Net's average accuracy, sensitivity, specificity and F1 score are 93.75%. Harangi et al. [26] in their study worked on a newly created segmentation dataset named APACS23 comprising of a total of 37,000 manually segmented images being applied to a combination of fully convolutional neural network (FCN). It is observed that the combination of FCN-8, FCN-16 and FCN-32 achieved an accuracy of 97.76%, Intersection of Union (IoU) value of 57.93% and Dice Score (DC) value of 73.36%.

Wubineh et al. [27] has proposed a model that improves segmentation efficiency by including the SPP bottleneck and atrous convolution into the SegNet framework, enabling the extraction of multiscale spatial information. The classification task uses the segmentation output as input. According to segmentation results, SPP-SegNet outperforms regular SegNet with 98.53% accuracy on the Pomeranian dataset. Additionally, it outperforms the typical SegNet, which achieves 90.95% accuracy on the SIPAKMED dataset, with 94.15%. For Pomeranian and SIPAKMED binary classification, SE-DenseNet201 obtains 93% and 99% accuracy respectively. Again, Liu et al. [28] in their work has implemented automated tumor contour segmentation that will be helpful in reducing the workload for radiation oncologists. T2 weighted images (T2WI) from 151 patients with cervical cancer (CC) are gathered for this study, and the dataset is now available publicly. To study their possible clinical uses, the researchers constructed a nnU-Net based automated model for segmenting the contours of cervical tumors. The nnU-Net-3D acquired a recall of 82.52%, an IoU of 70.96% and a DSC of 81.81%. According to this study, radiation oncologists' effort can be decreased by using the nnU-Net model to quickly and appropriately identify cervical cancer tumors with high repeatability.

The subsequent Table 1 illustrates a more comprehensive examination of the literature survey.

Table 1: Analysis of some research articles

Author/Year	Method	Dataset Used	mAP	Limitations
Geetha *et al*. 2020 [15]	ACM and ENN-TLBO	Kaggle dataset (DICOM images)	89.60	The study is employed on a small dataset containing only 250 images. This makes it challenging to extrapolate the findings to a larger dataset. Since cancer data is unbalanced; hence the suggested

				approach demonstrates good overall accuracy but low sensitivity.
Yuan *et al.* 2020 [16]	Segmentation model based on U-Net	Private dataset (Acetic and Iodine images)	Average accuracy value of 95.59% and DICE value of 61.64% in acetic images. Average accuracy value of 95.70% and DICE value of 63.80% in iodine images.	To improve the current model, a huge number of images, particularly high- definition images from colposcopes of different brands were needed. To access clinical value, a large-scale, prospective, multicenter clinical trial must be conducted.
Liu *et al*. 2021 [17]	Improved U-Net named SD-U-Net	Private dataset containing 52 cases of HSIL from colposcopy	Dice value of 77.31% and IoU value of 62.75%	This method uses huge trainable parameters thus taking more training time.
Zhao *et al*. 2021 [18]	Selective Edge Enhancement-based Nuclei Segmentation method (SEENS)	18 Whole slide images containing 395 cell nuclei	-	SEENS is referred to be a non-learning-based model, meaning it depends on manually hand-crafted features and mathematical operators. When confronted with new differences in cell morphology, staining patterns, this method might not generalize well, also not be suitable for deep learning techniques. The dataset is too small to evaluate a model.
Fan *et al.* 2022 [19]	Modified U-Net	Private dataset (6694 CT images)	Dice value of 87.60% and IoU value of 81.40%.	Additional data like MRI images can be included in future. Hence multi-model fusion can be applied to combine extra medical data from various devices.
Chatterjee *et al*. 2022 [20]	DRU-Net	SIPAKMED dataset (1850 images used)	90.66%	DRU-Net can be used in the future to segment overlapped cells in medical images.
Zhao *et al.* 2022 [21]	BC-Net, Mask RCNN, U-Net, U-Net++, Attention U-Net, AL-Net	New dataset introduced named CNSeg	Achieved highest AJI value of 71.32%, PQ value of 70.85%, HD value of 15.91% using AL-Net model.	Several deep learning models are proposed and evaluated in the study itself. On the CNSeg dataset, subsequently research has frequently outperformed the original models, indicating that the baseline models may not have fully utilized the potential of the new dataset or were not the state-of-the-art architectures at the time of publication.
Ji *et al.* 2023 [22]	Transfer learning ensemble of U-Net and U-Net++ utilizing Resnet34 and DenseNet121 as	Cx22 dataset	Dice value of 94.80% for cytoplasm segmentation and 75% for nucleus segmentation.	Because the study only used the Cx22 dataset and did not use any other external validation, the models' capacity to generalize across various imaging systems, labs or patient groups has not been confirmed.

	encoders			
Chowdhury *et al*. 2023 [23]	U-Net based novel model	Herlev	On the Herlev dataset, model has achieved 98.39% accuracy and recall of 98.97%	For actual preclinical application, our algorithms' performance has to be further refined. Furthermore, idea of using data resampling to balance the dataset and improving performance can be explored.
Rudiansyah *et al.* 2024 [24]	U-Net	Herlev	89.00% Accuracy and 78.00% IoU is achieved.	Although Herlev dataset is widely used as a benchmark dataset, it is smaller and less varied than contemporary large-scale datasets such as CNSeg.
Desiani *et al*. 2024 [25]	U-Net	Zenodo (private dataset)	93.00% Accuracy	Additional algorithms, including YOLO, RCNN, and others, can be used in future studies to identify additional cells, like cytoplasm, in pap smear pictures. This approach can be used to create automatic cervical cancer detection for the classification stage.
Harangi *et al*. 2024 [26]	FCN	APACS23	97.76% Accuracy, 57.93% IoU, and 73.36% DSC	The overall time needed to build the segmentation masks for a given image has greatly increased due to the structure of the proposed model.
Wubineh *et al.* 2025 [27]	SPP-SegNet	SIPAKMED (5 class)	94.15% Accuracy, 93.87% precision, 94.94% recall, 95.08% IoU	Using isolated single-cell images might not accurately represent the complexity of actual clinical samples. This is one of the study's limitations.
Liu *et al.* 2025 [28]	nnU-Net-3D	T2 weighted images (MRI dataset) from 151 patients.	Dice value of 81.81%, IoU of 70.96% and recall of 82.52%	The MRI dataset size is constrained. Hence the model's capacity to generalize is limited by a smaller dataset. The model might not be able to capture unusual tumor appearances that were underrepresented in the training set.

3. MATERIALS AND METHODS

This section talks about the tools and methods used to sort cervical cancer images. Following Figure 1 depicts the basic workflow of the proposed methodology for cervical cancer segmentation.

The experimental setup includes the following steps: (i) dataset setup – images and matching masks are put in other directories. Also, a checking is done so that the image-mask pairing is accurate, meaning same image numbers and similar filenames are present; and verification of the masks is done to check for proper class labels. (ii) data preprocessing –images and masks are resized to the desired scale. Then, the image's pixels are normalized typically to [0,1] and masks are transformed into integer class labels. (iii) data augmentation – color jittering (brightness, contrast, saturation, hue) is applied to input images. (iv) Scratch U-Net model – Encoder consisting of components of pre-trained backbones like ResNet, VGG, EfficientNet, MobileNet; Bottleneck – containing the deepest layer with highest channel count and acts as the bridge between encoder and decoder; Decoder - containing transposed convolutions or upsampling layers, skip connections and final 1x1 convolution for class prediction. (v) Loss Function Design – various standard loss functions like dice loss (works for boundary prediction and small objects) and focal

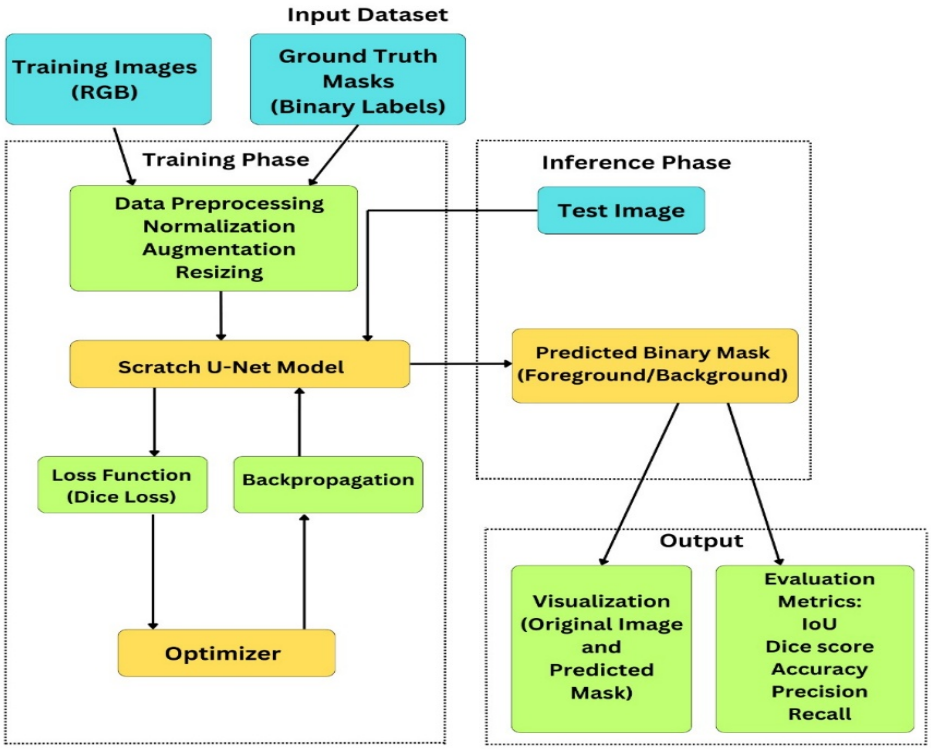

Figure 1: Proposed methodology workflow for segmenting cervical cancer cells

loss (handles extreme class imbalance) is applied. Dice loss is useful for datasets with extreme imbalance, when the foreground class (such as cancer) is significantly smaller than the background. (vi) Optimizer – The training strategy is applied by choosing a specific optimizer which modifies the model's parameters (weights and biases) to reduce the value of the loss function consequently improving the model performance. (vii) Test evaluation is performed based on various metrics like IoU, Dice, Accuracy etc. and also visual predictions versus ground truth is performed.

3.1 Dataset Description

The experimental investigation utilizes the publicly accessible CNSeg dataset [29]. With more than 1,24,000 annotated nuclei gathered from 1530 patients under various circumstances, CNSeg is the biggest publicly accessible resource for nuclear segmentation. This dataset is created especially to overcome the shortcomings of previous cervical cytology datasets, which includes low segmentation difficulties, single data sources and insufficient volumes. The dataset comprises patients with a range of ages and illnesses. Few distinct scanners are used to scan whole slide images (WSI), producing a variety of image styles and content. A total of 3010 image-mask pairs is available in this dataset. Following Table 2 shows the structure of train, test and validation datasets of the PatchSeg subset present in CNSeg dataset. For a thorough assessment, CNSeg is separated into three subsets: PatchSeg, a patch segmentation dataset containing images of nuclei taken under challenging circumstances. It includes tiny patch images that are cropped from WSIs under complicated circumstances such as overlapping nuclei, cytopathic heterogeneity and microbial infection. It is intended to assess how well segmentation procedures generalize.

Table 2: Distribution of images in train/valid/test

Category	Image-Mask Pair
Train	2107
Validation	451
Test	452

3.2 Dataset Visualization

The following Figure 2 includes tiny patch image samples that are clipped from whole slide images (WSIs) under challenging circumstances in order to assess how well segmentation techniques generalize.

Figure 2: Sample images from PatchSeg subset of CNSeg dataset

3.3 Splitting Data

Randomizing the data before splitting is crucial to ensure that each set is representative of the total dataset. The relative frequencies of various classes are preserved throughout all data splits using stratification. This avoids a scenario where a specific split might have extremely few or no samples of a minority class, which could result in biased model evaluation or training. Stratification is a technique used during data splitting (in training, validation and test sets) in segmentation tasks to guarantee that each resulting subset retails a similar proportion of classes as the original dataset. This lessens the potential for bias in the training and evaluation of the model and is crucial when working with imbalanced datasets. So, the dataset is split into three parts: 70 for training, 15 for testing and 15 for validation.

3.4 Data Augmentation and Image Processing

Despite the size of CNSeg having 124,000 annotated nuclei, augmentation can help models become more robust and learn invariancies. Using various factors such as cropping, color jittering and other techniques data augmentation is used to create new images from the original images. Data augmentation [30] makes the training dataset bigger and better in this way. Table 3 shows the several parameters that are taken into account when adding data.

Table 3: Data augmentation parameters

Parameters	Values
ColorJitter	hue=0.1, saturation=0.2, brightness=0.2, contrast=0.2

Again, Table 4 below shows the values we used to find the mean and standard deviation in our method. Normalization facilitates faster learning for neural networks, which in turn promotes faster convergence.

Table 4: Parameters used for normalizing the image

Parameters	Values
Standard deviation	[0.229,0.224,0.225]
mean	[0.485,0.456,0.406]

3.5 Building a Scratch Segmentation Model

A lightweight encoder-decoder framework is created especially for binary segmentation task using the components of YOLOv5 along with few other components in downscaled version in U-Net architecture. The proposed segmentation model combines the effective backbone of YOLOv5 with the decoder architecture of U-Net, which is improved by multi-scale deformable attention (MSDA) techniques. The encoder portion of this proposed segmentation model when applied as a normal classifier for cervical cell classification showed accurate performance [31]. This architectural congruence suggests that the underlying design can intrinsically extract generalizable high-quality multiscale representations that are successful for different task objectives, instead of being narrowly optimized for a single downstream task. The strength and flexibility of the backbone design itself is evidenced by the fact that the same encoder-neck architecture achieves good performance in both global image-level classification and fine-grained pixel-level segmentation independently. Following Tables 5, 6 and 7 contains the detailed architectural design of encoder, neck and decoder of the proposed YOLOv5++ model respectively.

Table 5: Encoder architecture

Layer Name	Type	Output Shape
RGB input image		(B,3,512,512)
Stem	Conv	(B, 16, 256, 256)
Stage1 - Conv	Conv	(B, 32, 128, 128)
Stage1 – C2f	C2f	(B, 32, 128, 128)
Stage2 – Conv	Conv	(B, 64, 64, 64)
Stage2 – C2f	C2f	(B, 64, 64, 64)
Stage3 – Conv	Conv	(B, 120, 32, 32)
Stage3 – C2f	C2f	(B, 120, 32, 32)
Stage4 – Conv	Conv	(B, 232, 16, 16)
Stage4 – C2f	C2f	(B, 232, 16, 16)
SPPF	SPPF	(B, 232, 16, 16)

Table 6: Neck Architecture

Layer Name	Type	Output Shape
Projection Stage2	Conv 1x1	(B, 232, 64, 64)
Projection Stage3	Conv 1x1	(B, 232, 32, 32)
MSDA	MultiScale Deformable Attention	(B, 232, 16, 16)
Fusion Conv	Conv	(B, 232, 16, 16)

Table 7: Decoder and Output architecture

Layer Name	Type	Output Shape
Decoder4 - Upsample	Upsample 2x	(B, 232, 32, 32)
Decoder4 – MSDA Gate	MSDA Attention Gate	(B, 120, 32, 32)
Decoder4 - Concat	Concatenate	(B, 352, 32, 32)
Decoder4- Conv	Conv	(B, 120, 32, 32)
Decoder4- C2f	C2f	(B, 120, 32, 32)
Decoder3 - Upsample	Upsample 2x	(B, 120, 64, 64)
Decoder3 – MSDA Gate	MSDA Attention Gate	(B, 64, 64, 64)
Decoder3 - Concat	Concatenate	(B, 184, 64, 64)
Decoder3- Conv	Conv	(B, 64, 64, 64)
Decoder3- C2f	C2f	(B, 64, 64, 64)
Decoder2 - Upsample	Upsample 2x	(B, 64, 128, 128)
Decoder2 – MSDA Gate	MSDA Attention Gate	(B, 32, 128, 128)
Decoder2 - Concat	Concatenate	(B, 96, 128, 128)
Decoder2- Conv	Conv	(B, 32, 128, 128)
Decoder2- C2f	C2f	(B, 32, 128, 128)
Decoder1 - Upsample	Upsample 2x	(B, 32, 256, 256)
Decoder1 – MSDA Gate	MSDA Attention Gate	(B, 16, 256, 256)
Decoder1 - Concat	Concatenate	(B, 48, 256, 256)
Decoder1- Conv	Conv	(B, 16, 256, 256)
Decoder1- C2f	C2f	(B, 16, 256, 256)
Output -> Final Upsample	Upsample 2x	(B, 16, 512, 512)
Output -> Conv1	Conv	(B, 8, 512, 512)
Output -> Conv2	Conv	(B, 4, 512, 512)
Output -> Segmentation Head	Conv2d	(B, num_classes, 512, 512)

Below Figure 3 displays the detailed architecture diagram of the proposed scratch YOLOv5++ model. Five steps are used by the encoder to extract hierarchical features. In stem stage, the input image is reduced to half resolution by the first 6x6 convolution with stride 2.

In stages 1 till 4, C2f modules of YOLOv8 are used. These are effective bottleneck blocks with split concatenation and are used after a regular convolution to downsample by 2x. Following Stage 4, spatial pyramid pooling fast (SPPF) rapidly captures multi-scale global contextual information by applying cascaded max pooling technique for multi-scale pooling operation. The neck improves the bottleneck features by utilizing multi-scale deformable attention. The neck uses the output of the bottleneck as the query focuses on characteristics from Stages 2, 3 and 4. It uses weighted aggregation to combine data from multiple scales. It also incorporates a fusion convolution and residual connection. Hence this enables the model to choose concentrate on pertinent information ay various spatial resolutions. Each of the four progressive upsampling blocks that make up the decoder includes bilinear upsampling (2x) to improve spatial resolution. MSDA attention gate uses the upsampled features as a gating signal to apply multi-scale deformable attention to the skip connection. This process suppresses unimportant information while highlighting key aspects of the skip connection. After this, concatenation combines upsampled characteristics

with the gated skip connection. The convolution and C2f refines the output by processing the combined features. The output head consists of final upsampling a two-dimensional bilinear upsampling using two channel-reducing progressive convolutions. Lastly the segmentation head consists of the final per pixel class predictions generated from 1x1 convolution.

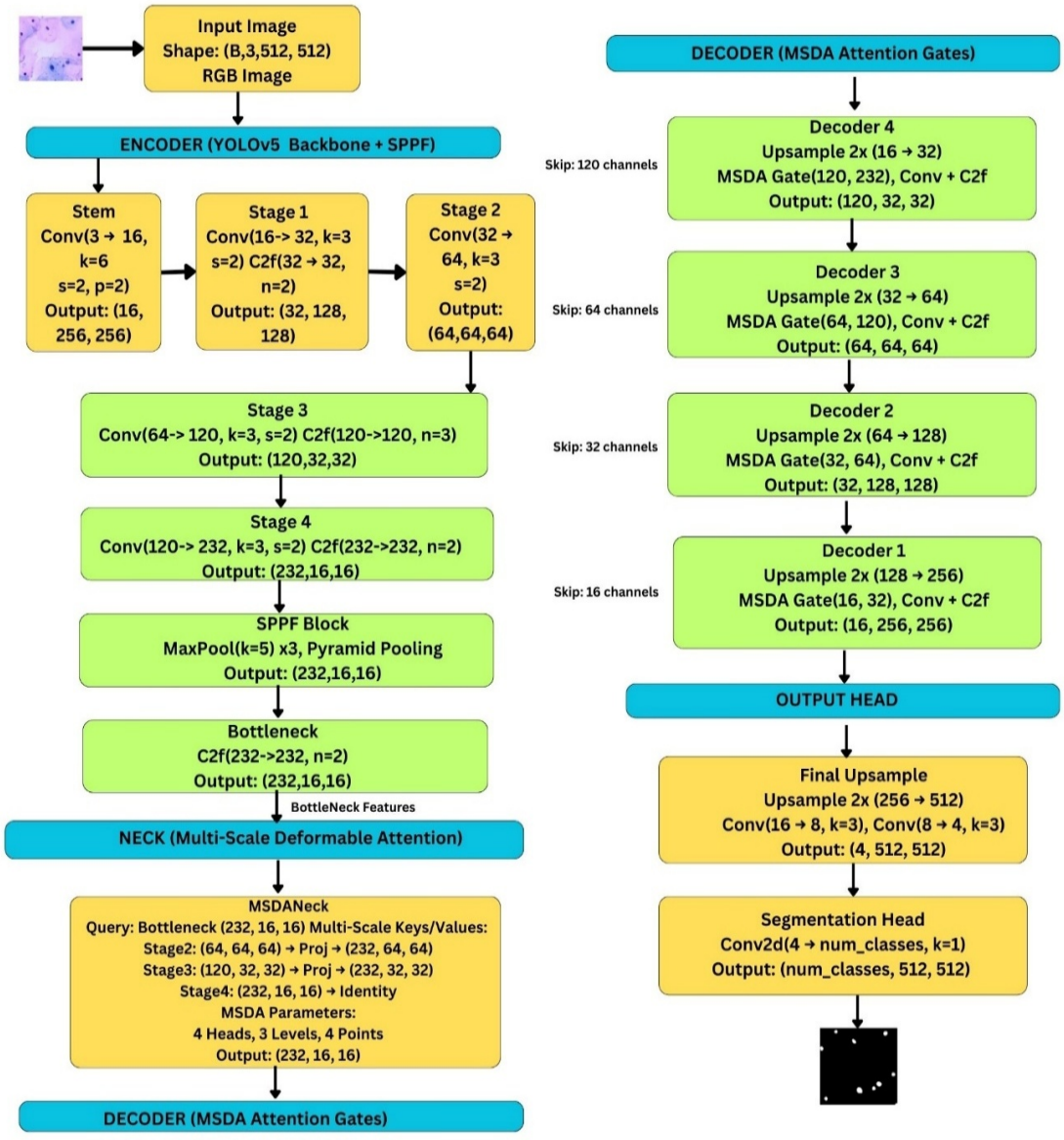

Figure 3: Architectural schematic of the proposed scratch U-Net model for cervical cancer image segmentation: Input image and Encoder, Neck, Decoder and Output Head

3.6 Software and Hardware Used

The experiments have been carried out in Pytorch (2.5.1) and Torchvision (0.20.1) on a Google Colab Notebook on an NVIDIA A100 GPU. The device has a 16GB RAM and an i5 11th generation processor.

3.7 Model Parameters

Below Table 8 lists all of the model parameters that are taken into consideration. The proposed model is executed for 50 epochs using learning rate value of 0.0008. We have used well-known loss functions such as dice loss and focal loss in image

segmentation. Dice loss is utilized in challenges involving image semantic segmentation.

Table 8. Hyperparameters used in the proposed model

Parameters	Values
Learning Rate	0.0008
Loss Function	Dice Loss, Focal Loss
Batch Size	8
Epoch Number	50
Size of Image	512 X 512 X 3
Optimizer	Adam
Scheduler	CosineAnnealingWarmRestarts

It calculates the overlap between segmentation masks that are expected and those that are ground truth. Whereas focal loss reduces the weight of simple instances to address the severe foreground-background class imbalance. Together they balance pixel-accuracy and spatial overlap. CosineAnnealingWarmRestarts scheduler is used in the experimental setup to increase the neural network training speed and efficiency. For better performance optimal batch size of 8 is considered.

4. RESULTS AND DISCUSSION

This part talks about the analysis and observations that come from the experimental results. The most often employed and acknowledged evaluation measures in segmentation task research are Intersection Over Union (IoU), Dice Similarity Coefficient (DSC), accuracy, precision, recall and F1-score.

4.1 Evaluation Parameters Used in Experimental Study

In a binary segmentation task, each pixel in an image is classified into one of two groups, usually Background (negative class) or Foreground (positive class). The following four outcomes are used to calculate the metrics accuracy, precision, recall and F1 score; which are often summarized on a per-pixel basis in a confusion matrix.

TP (True Positive) – pixels that are correctly classified as foreground.

TN (True Negative) – pixels that are accurately projected to be background.

FP (False Positive) – pixels mistakenly identified as foreground (in reality, actual class is background).

FN (False Negative) – pixels mistakenly identified as background (in reality, actual class is foreground).

Table 9 displays the equations for all the performance evaluation metrics used in evaluating our proposed model. The very tiny constant 1e-8 is added to the denominator. It is a standard procedure in numerical computing to avoid division by zero. The overlap between two bounding boxes or segmentation masks is measured using the IoU formula, a fundamental metric used in object detection and segmentation. The IoU value ranges between 0 and 1. An IoU value of 0 indicates no overlap between the predicted and ground truth regions, whereas 1 represents a complete overlap between the predicted and ground truth regions. The intersection area divided by the total number of elements in both sets (ground truth mask and the predicted mask) yields the Dice Similarity Index. The two main measurements used are dice and IoU, with dice providing a more accurate measurement for smaller targets.

Table 9: Evaluation parameters used in segmentation task

Performance Metrics	Formula
Accuracy	TP+TN/(TP+TN+FP+FN+1e-8)
Precision	TP/TP+FP+1e-8
Recall	TP/TP+FN+1e-8
F1-score	(2*P*R)/(P+R+1e-8)
IoU	Area of Intersection/Area of Union
DSC	(2*Number of pixels present in both ground truth and the model's output)/Total number of pixels in both shapes combined

4.2 Ablation Study

The contribution of various architectural elements to the YOLOv5 based U-Net model for segmentation task is methodologically assessed in this ablation study. Below Table 10 contains the sequential addition of architectural enhancements. The final configuration consisting the last row of the table combines all of the improvements (C2f, SPPF, MSDA, autocast and Adam optimizer).

It has the most balanced performance and the greatest IoU value of 69.68%. Remarkably, the Dice score rises to 82.13% demonstrating strong positive occurrence detection, improving both boundary precision and regional coverage. As our work will aid the work of pathologists and is not a real-time application, hence in such situations, where precision is crucial, the 2.46M parameter model offers improved segmentation quality, making this slight increase in model complexity justifiable.

Table 10. Ablation study of various components used in YOLOv5++ model

Component	Accuracy (%)	Dice (%)	IoU (%)	Params (M)
YOLOv5 baseline	98.21	81.51	68.79	9.82
YOLOv5 + C2f + multi-scale deformable attention + autocast	98.25	81.68	69.03	2.33
YOLOv5 + C2f + SPPF + multi-scale deformable attention + autocast	**98.33**	**82.13**	**69.68**	2.46

4.3 Proposed Model Results

Figure 4 shows the confusion matrix where the majority of the model's predictions fall on the diagonal, indicating excellent overall accuracy.

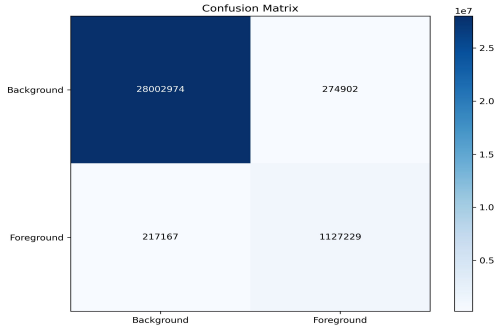

Figure 4: Confusion Matrix of YOLOv5++ model

The models perform very well in the background class with good specificity. Whereas Figure 5 displays the ROC and PR curves of our proposed model. The left side of this plot shows true positive rate and false positive rate at various categorization thresholds. The ROC-AUC of 0.99 indicates that the suggested model has near-perfect discriminative capability, accurately differentiating between positive and negative classes for almost all classification thresholds. Again, the right side of Figure 5 plots the precision vs recall graph. The model's value PR-AUC of 0.89 suggests great precision over the majority of recall values. It is especially helpful for datasets that are unbalanced. The blue curve shows the trade-off between precision and recall.

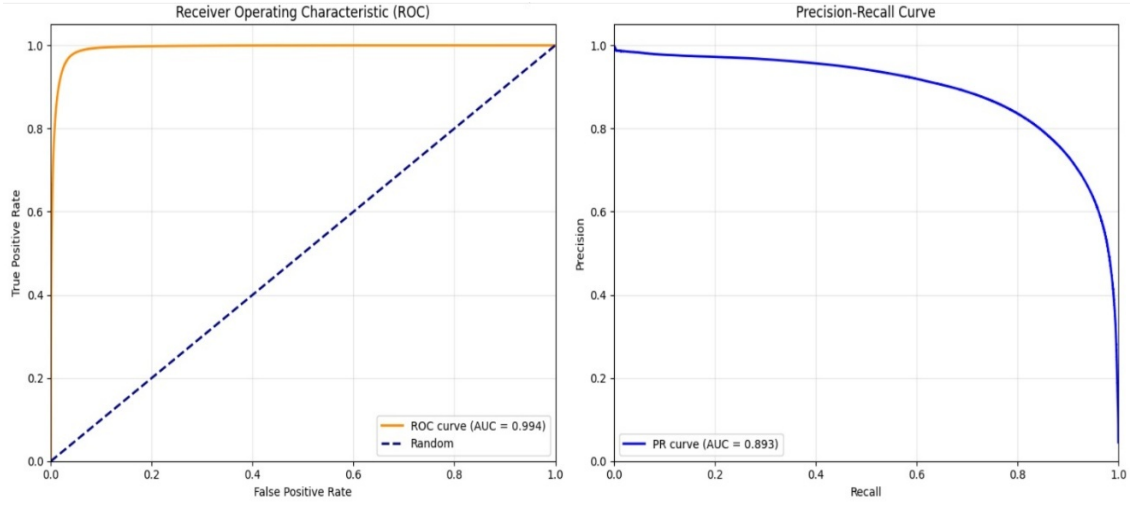

Figure 5: ROC-PR curve of YOLOv5++ model

The YOLOv5++ segmentation model performs well across all five input cervical cancer images when compared to the ground truth and expected segmentation masks. Below Figure 6 shows the predicted masks closely match the ground truth annotations, effectively identifying and localizing individual cell nuclei with great spatial accuracy. The anticipated and actual segmentations have good to outstanding overlap, as indicated by the intersection over union (IoU) scores, which range from 0.63 to 0.87. The model exhibits significantly lower but still significant performance on more

difficult situations with overlapping cells (IoU value 0.63) and performs very well on images with well-separated nuclei (IoU value 0.87). Ground truth masks and the predicted masks visually preserve similar nucleus sizes, shapes and locations; small differences mostly occur near object boundaries or in densely populated cellular areas. The model has learnt reliable features for nucleus detection and segmentation in histopathology images based on its consistent performance across different staining intensity and cellular densities. A comparative analysis of various CNN encoder architectures for U-Net based semantic segmentation is presented in Table 11. When compared to baseline implementations utilizing well-known deep learning architectures, the suggested YOLOv5++ model performs best across a variety of evaluation metrics implemented. With an accuracy of 98.33%, the suggested scratch segmentation model outperforms all baseline architectures. This is significantly better than MobileNetv2 (98.05%), the second-best performance and far better than ViT (91.82%), the worst performer. With an IoU of 69.68% and a Dice coefficient of 82.13%, the suggested YOLOv5++ model outperforms all other tested designs in terms of spatial segmentation quality. Because they directly assess the quality of boundary delineation and pixel-wise categorization, these measures are very important for segmentation tasks. These findings show that the proposed model maintains high pixel-wise accuracy and spatial overlap with ground truth annotations while achieving greater segmentation quality through an enhanced balance between precision and recall.

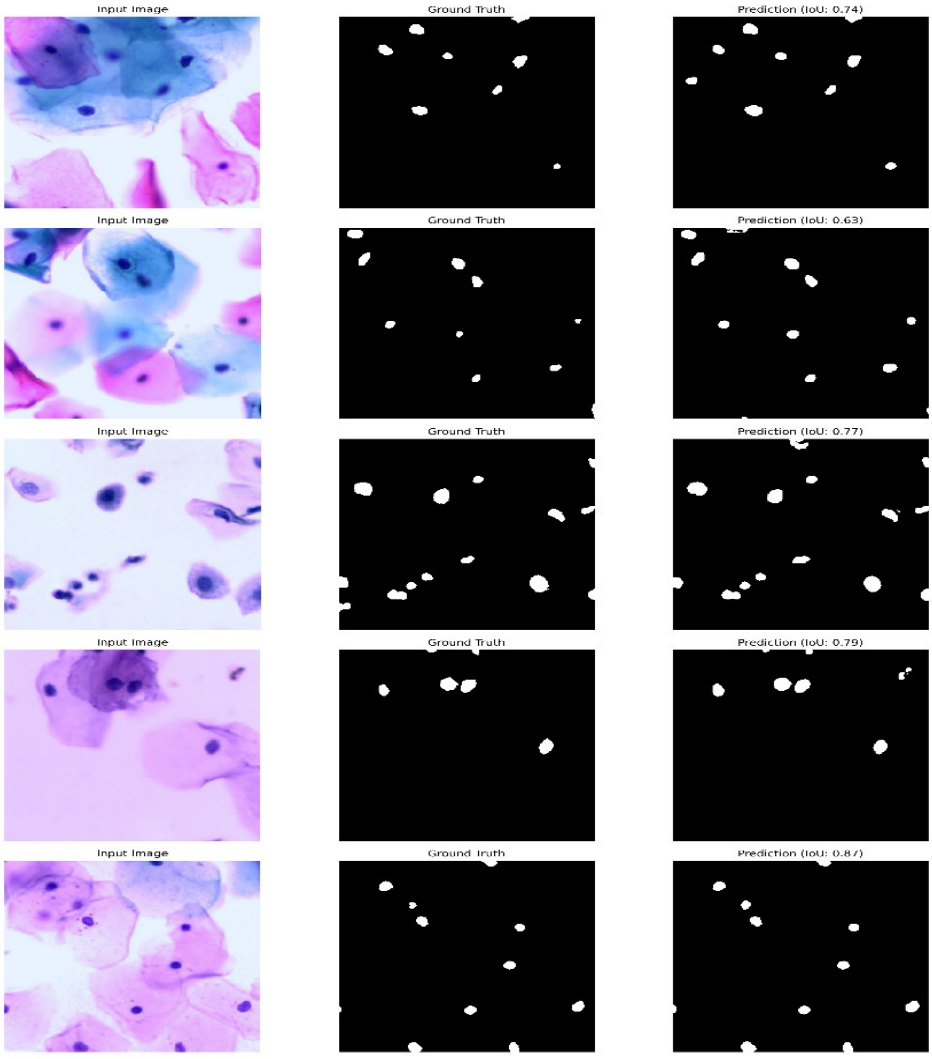

Figure 6. Segmentation visualizations using YOLOv5++ model

Table 11. Comparative analysis of our proposed model and various scratch CNN encoders

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	IoU (%)	Dice (%)
CNN encoder	0.9802	0.7253	0.9062	0.8057	0.6747	0.8114
VGG16 encoder	0.9813	0.7491	0.8729	0.8063	0.6755	0.8149
MobileNetv2 encoder	0.9805	0.7618	0.8312	0.7950	0.6597	0.8040
ResNet34 encoder	0.9803	0.7264	0.9074	0.8069	0.6762	0.8131
ViT encoder	0.9182	0.3193	0.7083	0.4402	0.2822	0.4492
YOLOv5 encoder	0.9821	0.7655	0.8779	0.8123	0.6879	0.8177
ViT+YOLOv5 encoder	0.9655	0.5809	0.8613	0.6939	0.5364	0.6939
Proposed YOLOv5++	**0.9833**	**0.7969**	0.8473	**0.8213**	**0.6968**	**0.8213**

The following Figure 7 shows how well a segmentation model performs on six important evaluation metrics. At 98.33% the model's accuracy is remarkably high, suggesting a strong overall predictive capacity.

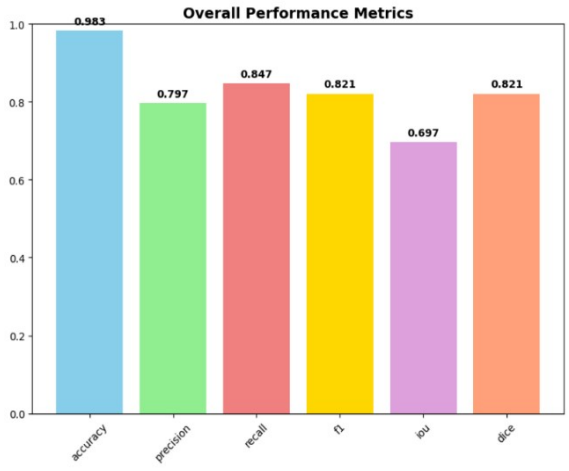

Figure 7: Performance summary of our proposed YOLOv5++ segmentation model

According to the comparison analysis of Table 12, our suggested model outperforms well-known encoder architectures in terms of computational efficiency. Among all analyzed architectures, the suggested YOLOv5++ model achieves the most lightweight configuration with 24,64,786 trainable parameters and 1.965G FLOPs. When compared to the VGG16 encoder (2,58,67,010 parameters) and the ViT encoder (9,69,84,290 parameters), the suggested model reduces trainable parameters by roughly 90.47% and 97.46% respectively. With just 2.465M parameters and 1.965G FLOPs, roughly 75% fewer parameters than MobileNetv2 and 77% fewer FLOPs than YOLOv5, this YOLOv5++ model achieves an exceptional efficiency. Because of this, it is the lightest and most computationally effective choice. This significant drop in FLOPs and parameter count suggests that the proposed paradigm is ideal for implementation in situations with limited resources, such as embedded systems and mobile devices as well.

Table 12. Comparative analysis on the parameters and flops

Model	Trainable Parameters	GFLOPs
CNN encoder	31.037M	109.484G
VGG16 encoder	25.867M	28.347G
MobileNetv2 encoder	9.517M	9.843G
ResNet34 encoder	24.426M	17.321G
ViT encoder	96.984M	85.684G
YOLOv5 encoder	9.828M	8.540G
ViT+YOLOv5 encoder	105.387M	166.024G
Proposed YOLOv5++	**2.465M**	**1.965G**

4.4 Explainability and Interpretability

Figure 8 presents a Grad-CAM based interpretability analysis of the proposed model using the final decoder layer (Decoder1 C2f). The input image and corresponding ground truth segmentation are presented with the Grad-CAM heatmap and the final forecast. The warm colors (red/orange) correspond to the areas having the highest influence on the decision of the model while the cooler colors (blue/cyan) correspond to

smaller contributions. The heatmap localizes mainly around the nuclei aligning with the ground truth identified by experts, which suggests that the model learns the morphologies associated with the clinical outcome, rather than artifacts in the background. The final prediction likewise agrees well with the ground truth, suggesting that the highlighted regions directly contribute to accurate nuclear segmentation and further confirming the interpretability and dependability of the model.

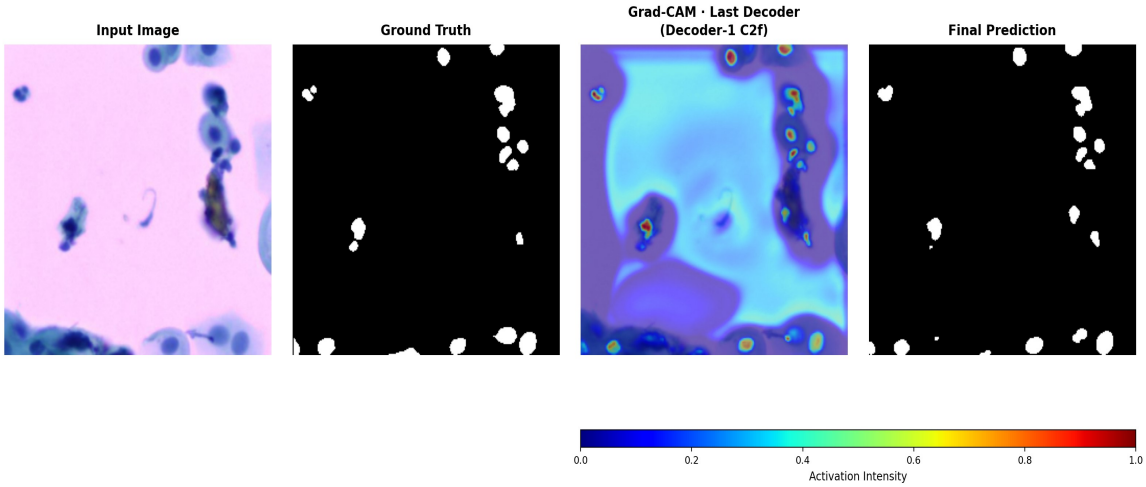

Figure 8: Grad-CAM analysis on a sample input image

4.5 Statistical Validation of Model Comparisons

Following Figure 9 and Figure 10 shows the output for the Cohen's d test.

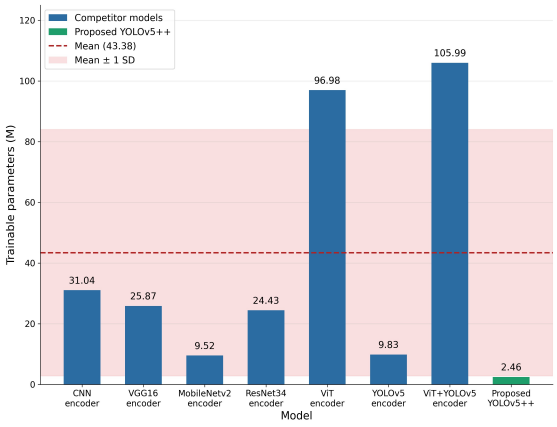

Figure 9: Effect size analysis using Cohen's d test on total parameters (Cohen's d = 1.055)

As we have one proposed model versus a set of 7 competitors, Cohen's d statistical test is the suitable test to examine the performance of YOLOv5++ model across all other models. We calculate d = (mean of other models – YOLOv5++) / (Standard Deviation of others). A positive d suggests YOLOv5++ is superior. Regarding the proposed YOLOv5++ model, the reduction in trainable parameters (d=1.055) and computing cost (d=1.002) is very large compared to the comparison models. This means that the efficiency benefits discovered are not only statistically significant but also of practical importance in terms of size.

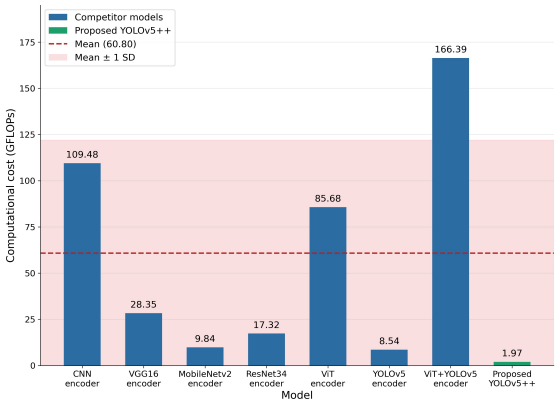

Figure 10: Effect size analysis using Cohen's d test on GFLOPs (Cohen's d = 1.002)

We have also performed Wilcoxon-Signed Rank test to understand how well our model is performing across all the various other deep learning models for segmentation work. The largest absolute difference comes from the ViT+YOLOv5 encoder (102.927M) and the ViT encoder (94.524M), which have the highest Wilcoxon rank values of 7 and 6

respectively, showing their significantly higher parameter burden compared to the proposed model. The CNN encoder (28.577M, rank 5), VGG16 encoder (23.407M, rank 4) and ResNet34 encoder (21.966M, rank 3) follow in declining order, with atleast an order of magnitude more parameters than the proposed YOLOv5++. Even the relatively small YOLOv5 encoder (7.368M, rank 2) and MobileNetv2 encoder (7.057M, rank 1) designs built for computational efficiency require around 3 times more trainable parameters. Importantly, the seven absolute differences are all the strictly positive, which proves that the proposed YOLOv5++ has the lowest trainable parameter count (2.46M) across all the competitors without exception. The Wilcoxon signed-rank test gives W+= 28 and W-= 0. The test statistic is W = 28.0 and p = 0.0078. This reduction in parameters is statistically significant at the 1% significance level. The monotonically decreasing rank line from rank 7 to rank 1 as shown in Figure 11 further confirms a steady and ordered trend of improvement, with no outlier or reversal in the ranking structure.

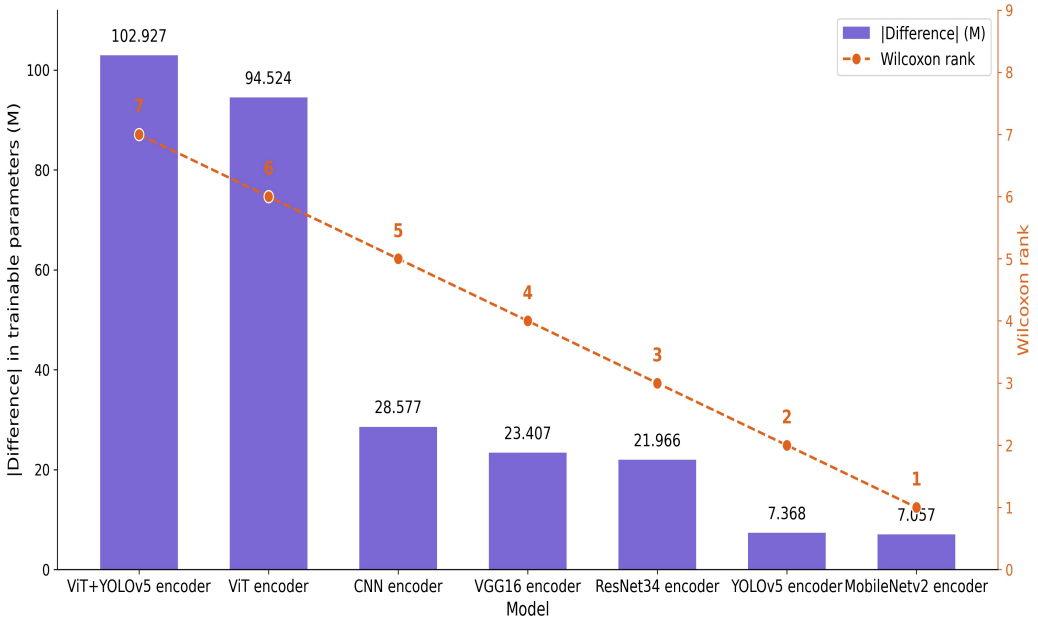

Figure 11: Wilcoxon signed-rank test: absolute differences and rank (parameters, sorted by magnitude)

These results indicate that the proposed YOLOv5++ outperforms all competing architectures in terms of parameter efficiency with a statistically validated and practically significant margin. All the seven absolute differences are positive without a single exception demonstrating that the proposed YOLOv5++ attains the lowest computational cost among the whole comparison set. The Wilcoxon signed-rank test produces a flawless positive rank sum of W+= 28 and W-= 0, resulting in a test statistic of W = 28.0 and p = 0.0078, so affirming that the detected computational advantage is statistically significant at the 1% level and is not due to random variation. The consistently decreasing rank line from rank 7 to rank 1 illustrates an uninterrupted order of magnitude across all competitors hence enhancing the reliability of the nonparametric test results. Figure 12 illustrates that the proposed YOLOv5++ provides the most computationally efficient inference pipeline among all evaluated designs. With merely 1.965 GFLOPs, it attains a reduction of upto 98.80% compared to the most expensive competitor ViT+YOLOv5, while preserving a statistically significant superiority over specialized lightweight designs like MobileNetv2 and YOLOv5. The suggested approach is well-suited for real-time, low latency deployment in mobile platforms, edge inference hardware and embedded vision systems with limited computational resources.

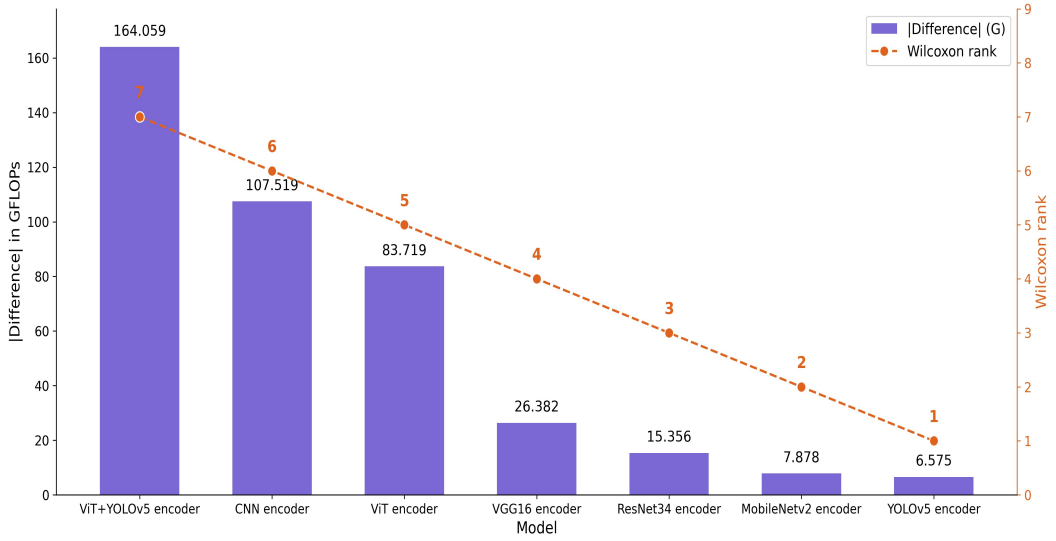

Figure 12: Wilcoxon signed-rank test: absolute differences and rank (GFLOPs, sorted by magnitude)

5. CONCLUSION AND FUTURE WORK

With consistently high scores on all measures, the model performs exceptionally well. Accuracy is the most notable metric, while IoU is the main area of potential improvement. The model appears to be a good fit for real-world semantic segmentation applications based on its balanced performance in precision, recall, F1-score and dice coefficient. According to experimental results, our model operates at a significantly low FLOP count value of 1.965G, indicating a high accuracy-efficiency trade-off and beats best methods for accuracy, dice and IoU. Future work will focus on adding more imaging modalities and assess it on a wider variety of medical datasets. By creating few-shot learning procedures or domain adaptation techniques, the model can be checked to function on untested datasets, improving its clinical usefulness. The suggested framework demonstrated promising segmentation performance, it suffers from a significant constraint in delineating nuclear peripheries, especially in the situation of overlapping or densely clustered nuclei. This limitation can be alleviated by including multi-scale feature fusion methods in future work.

ACKNOWLEDGEMENTS

The author extends deep gratitude to Dr. Rajdeep Chatterjee and Dr. Mahendra Kumar Gourisaria for their constant and unwavering support, without which this study would not have been possible.

REFERENCES:

[1] Who Facts. Available: https://www.who.int/news-room/fact-sheets/detail/cervical-cancer. Accessed on: 15/08/2025

[2] G. Turashvili, HPV associated cervical squamous cell carcinoma. Available: https//www.pathologyoutlines.com/topic/cervixSCC.html. Accessed on: 17/07/2025

[3] S.F. Ahmed, M.S.B. Alam, M. Hassan, M.R. Rozbu, T. Ishtiak, N. Rafa, M. Mofijur, A.B.M. Shawkat Ali and A.H. Gandomi. "Deep learning modelling techniques: current progress, applications, advantages, and challenges", *Artificial Intelligence Review*, vol. 56, no. 11, 2023, pp.13521-13617.

[4] H. N. Dai, R. C. W. Wong, H. Wang, Z. Zheng, & A. V. Vasilakos, "Big data analytics for large-scale wireless networks: Challenges and opportunities", *ACM Computing Surveys (CSUR)*, vol. 52, no. 5, 2019, pp. 1-36.

[5] Y. LeCun, Y. Bengio & G. Hinton, "Deep learning", *nature*, vol. 521, no. 7553, 2015, pp. 436-444.

[6] J. Lu, E. Song, A. Ghoneim, & M. Alrashoud, "Machine learning for assisting cervical cancer diagnosis: An ensemble approach", *Future Generation Computer Systems*, vol. 106, 2020, pp. 199-205.

[7] J. Long, E. Shelhamer & T. Darrell, "Fully convolutional networks for semantic segmentation", *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431-3440.

[8] O. Ronneberger, P. Fischer & T. Brox, "U-net: Convolutional networks for biomedical image segmentation", *In International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234-241.

[9] X. Pan, L. Li, D. Yang, Y. He, Z. Liu & H. Yang, "An accurate nuclei segmentation algorithm in pathological image based on deep semantic network", *IEEE Access*, vol. 7, 2019, pp. 110674-110686.

[10] H. Jung, B. Lodhi & J. Kang, "An automatic nuclei segmentation method based on deep convolutional neural networks for histopathology images", *BMC Biomedical Engineering*, vol. 1, no. 1, 2019, p. 24.

[11] J. Sikder, U. K. Das & R. J. Chakma, "Supervised learning-based cancer detection", *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, 2021.

[12] B. Z. Wubineh, A. Rusiecki & K. Halawa, "Segmentation and classification techniques for Pap smear images in detecting cervical cancer: a systematic review", *IEEE Access*, vol. 12, 2024, pp. 118195-118213.

[13] M. S. Devi, S. Priya, U. Desai, B. Acharya & F. Moreira, "Enhanced Histopathologic Image Analysis for Mouth Cancer Classification Using Morphological Reconstruction and UNet", *Journal of Imaging Informatics in Medicine*, 2025, pp. 1-24.

[14] B. Li, L. Chen, C. Sun, J. Wang, S. Ma, H. Xu, ... & L. Zhang", Leveraging deep learning for early detection of cervical cancer and dysplasia in China using U-NET++ and RepVGG networks", *Frontiers in Oncology*, vol. 15, 2025, p. 1624111.

[15] S. Geetha & S. Suganya, "Automatic Detection of Cervical Cancer Using UKF and ACM-AHP", *Int. J. Adv. Res. Eng. Technol*, vol. 11, no. 8, 2020, pp. 285-295.

[16] C. Yuan, Y. Yao, B. Cheng, Y. Cheng, Y. Li, Y. Li, ... & W. Lu, "The application of deep learning based diagnostic system to cervical squamous intraepithelial lesions recognition in colposcopy images", *Scientific reports*, vol. 10, no. 1, 2020, p.11639.

[17] J. Liu, Q. Chen, J. Fan & Y. Wu, "HSIL colposcopy image segmentation using improved U-Net", *In 2021 36th Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, 2021, pp. 891-897.

[18] M. Zhao, H. Wang, Y. Han, X. Wang, H. N. Dai, X. Sun, ... & M. Pedersen, "Seens: Nuclei segmentation in pap smear images with selective edge enhancement", *Future Generation Computer Systems*, vol. 114, 2021, pp. 185-194.

[19] Y. Fan, Z. Tao, J. Lin & H. Chen, "An encoder-decoder network for automatic clinical target volume target segmentation of cervical cancer in CT images", *International Journal of Crowd Science*, vol. 6, no. 3, 2022, pp. 111-116.

[20] P. Chatterjee, S. R. Dutta & M. Rafeeq, "Comparison of segmentation of CU-Net, DRU-Net and DU-Net on Cervical Cancer Data Set", *In 2022 8th International Conference on Smart Structures and Systems (ICSSS)*, 2022, pp. 01-05.

[21] J. Zhao, Y. J. He, S. H. Zhou, J. Qin & Y. N. Xie, "CNSeg: a dataset for cervical nuclear segmentation", *Computer Methods and Programs in Biomedicine*, 2023, pp. 241:107732.

[22] J. Ji, W. Zhang, Y. Dong, R. Lin, Y. Geng & L. Hong, "Automated cervical cell segmentation using deep ensemble learning", *BMC medical imaging*, vol. 23, no. 1, 2023, p.137.

[23] Chowdary, G. J., G, S., M, P., & Yogarajah, P., "Nucleus segmentation and classification using residual SE-UNet and feature concatenation approach in cervical cytopathology cell images", *Technology in Cancer Research & Treatment*, vol. 22, 2023, p. 15330338221134833.

[24] R. Rudiansyah, L. Iryani, L.I. Kesuma, P. Sari & A. Alamsyah, "Combination of image enhancement and u-net architecture for cervical cell semantic segmentation", *Journal of Informatics and Telecommunication Engineering*, vol. 7, no. 2, 2024, pp. 575-586.

[25] A. Desiani, Y. Utama, M. Arhami, A. K. Affandi, M. A. Sasongko & I. Ramayanti, "Simple Data Augmentation and U-Net CNN for Neclui Binary Segmentation on Pap Smear Images", *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 6, no. 3, 2024, pp. 264-275.

[26] B. Harangi, G. Bogacsovics, J. Toth, I. Kovacs, E. Dani & A. Hajdu, "Pixel-wise segmentation of cells in digitized Pap smear images", *Scientific Data*, vol. 11, no. 1, 2024, p. 733.

[27] B. Z. Wubineh, A. Rusiecki & K. Halawa, "SPP-SegNet and SE-DenseNet201: a dual-model approach for cervical cell segmentation and classification", *Cancers*, vol. 17, no. 13, 2025, p. 2177.

[28] G. Liu, Y. Tang, L. Guo, L. Zhang, Q. Zheng, H. Hui ... & Y. Zhou, "An open MRI dataset for cervical cancer with tumor segmentation using nnU-net", *Signal, Image and Video Processing*, vol. 19, no. 9, 2025, p. 692.

[29] Dataset URL. Available online: https://www.kaggle.com/datasets/zhaojing0522/cervical-nucleus-segmentation. Accessed on: 15/08/2025

[30] C. Shorten & T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning", *Journal of big data*, vol. 6, 2019, no. 1, pp. 1-48.

[31] J. Mondal, R. Chatterjee and M.K. Gourisaria, "Multiscale Deformable Attention Enhanced YOLO Framework for Cervical Cancer Cell Classification", *In 2026 7th International Conference on Computational Intelligence and Networks (CINE),* 2026, pp. 1-7.