Cervix-Seg-Net: A Context-aware Attention mechanism for Cervical Lesions Segmentation using Colposcopic images

Smith K. Khare[1,2] SMKH@MMMI.SDU.DK
[1] *The Applied AI and Data Science Unit, Mærsk Mc-Kinney Møller Institute, Faculty of Engineering, University of Southern Denmark, Denmark*
[2] *The Centre for Clinical Artificial Intelligence, Odense University Hospital, Denmark*

Berit Bargum Booth[3] BERIT.BOOTH@AUH.RM.DK
[3] *The Centre for Department of Gynecology and Obstetrics, Odense University Hospital, Denmark University Hospital, Denmark*

Victoria Blanes-Vidal[1] VBV@MMMI.SDU.DK

Lone Kjeld Petersen[3,4] LONE.KJELD.PETERSEN@RSYD.DK
[3] *The Centre for Research Unit for Gynecology and Obstetrics (Odense), Odense University Hospital, Denmark*
[4] *The Centre for Department for Clinical Research at University of Southern Denmark, Denmark*

Esmaeil S. Nadimi[1] ESI@MMMI.SDU.DK

Editors: Under Review for MIDL 2026

Abstract

Colposcopic examination is the diagnostic procedure in women who are positive in the cervical cancer (CC screening programme. Currently, a gynecologist expert performs colposcopy and manually marks high-risk cervical lesions. This process can be time-consuming and prone to inter- and intra-observer variations. As a result, there is an urgent need for an artificial intelligence-based solution for automated mapping of high-risk cervical lesions for effective CC screening. However, the current techniques are limited to classifying cervical lesions and identifying regions of interest. Therefore, this paper explores an effective and accurate segmentation of high-risk cervical lesions using a novel Cervix-Seg-Net model. The proposed model comprised a hybrid combination of an EfficientNet-based encoder to extract multi-scale, high-resolution contextual representations, pyramid scene parsing to capture global and regional context, attention mechanisms for adaptive weighting, and refined skip connections for noise reduction. This novel architecture addresses the challenges of limited contextual awareness and boundary ambiguity to provide precise delineation of high-risk cervical lesions. The Cervix-Seg-Net evaluation yielded dice scores of 0.847, 0.811, and 0.748 for training, validation, and test sets, respectively, using holdout and cross-validation techniques on a unique dataset with exact mapping of the cervix. The results indicate the model's robust performance and effective lesion localization, making it a potential candidate for real-world clinical integration and decision support.

Keywords: Digital colposcopy, deep learning, semantic segmentation, and cervical lesion.

1. Introduction

Cervical cancer (CC) remains the second most common malignancy among women of reproductive age worldwide, both in incidence and mortality. This burden is disproportionately higher in countries with lower Human Development Index (HDI) scores (Sung et al., 2021; Arbyn et al., 2020). CC follows a well-characterized and stepwise precancerous trajectory. It progresses from a human papillomavirus (HPV) infection to a low-grade squamous intraepithelial lesion and subsequently to a high-grade squamous intraepithelial lesion (Sung et al., 2021). Because of clinical advancements and early CC screening, it is now recognized as one of the most preventable and treatable malignancies (Wuerthner and Avila-Wallace, 2016). Colposcopic biopsy in diagnosing cervical lesions among screen positive women is considered the gold standard in diagnosing cervical lesions. During colposcopic examination, experienced gynecologists manually identify and highlight the high-risk cervical lesions by applying acetic acid solution to the cervix. However, colposcopy examination is subjective, depends on expert experience, and has low sensitivity, with high inter- and intra-observer variations (Li et al., 2023). Previous studies have shown that multiple biopsies improve overall sensitivity (Wentzensen et al., 2015; Gage et al., 2006; Booth et al., 2021). However, colposcopic biopsy leads to post-procedural effects, such as transient pelvic discomfort, minor vaginal bleeding, and short-term serosanguinous discharge as well as increased work load for analysis of multiple biopsies. Therefore, there is a need for an automated solution to accurately recommend biopsy sites corresponding to high-risk cervical lesions, thereby improving diagnostic accuracy while minimising patient discomfort. Recent advancements in medical artificial intelligence (AI) have enabled computational tools to assist colposcopists and gynecologists in identifying cervical lesions more precisely and effectively. However, existing convolutional neural network (CNN)- based models remain limited to image classification and cervix region identification. These models fall short in segmentation tasks. This motivates us to develop an effective segmentation model to accurately segment high-risk cervical lesions and evaluate their relative size.

2. Literature survey

Various machine learning and deep learning models for different segmentation tasks in CC have been developed using colposcopy and cervigram techniques. Table 1 summarizes the existing studies for various segmentation tasks.

Current state-of-the-art models, such as UNET with skip connections, can transfer unnecessary background data and struggle with long-range dependencies, potentially compromising robustness in colposcopy imaging (Chibuike and Yang, 2024). Models with Atrous Spatial Pyramid Pooling (ASPP) capture multiscale context but produce gridding artifacts and equally scaled weights, limiting their effectiveness with variable cervical lesions (Lian et al., 2021). FPN improves multiscale fusion, but its fixed approach weakens subtle lesion edges, while transformer-based models are computationally expensive and require extensive labeled data (Xu et al., 2021; Chibuike and Yang, 2024). Furthermore, existing studies are mainly limited to cervix segmentation but lack detailed lesion-level segmentation (Park et al., 2022; He et al., 2024; Dash et al., 2023; Bai et al., 2018; Bolaños Semanate et al., 2024). These limitations demand a need for a framework that can:

Table 1: Summary of studies for segmentation tasks in CC using different modalities.

Ref.	Source	Subjects/ images	Model	Description	Performance
(Yu et al., 2022b)	Cervigram	906 subjects	Lesion segmentation	Multi-phased model with region of interest identification using Fast-RCNN to focus on main region of cervix, followed by lesion segmentation using efficientNet-ASPP module.	DSC: 0.7431
(P and M, 2018)	–	–	Classification and segmentation	A grey level co-occurrence matrix and local binary patterns for classification using neural network followed by morphological operation to segment lesions from abnormal cervix images.	Recall: 97.42
(Park et al., 2022)	Colposcopy	270 images	Cervix segmentation	Encoder-Weighted learning in Graphcut W-Net model for cervical region segmentation from colposcopic images	DSC: 0.7100
(He et al., 2024)	Colposcopy		Cervix segmentation	A refining bottleneck based shuffle attention Residual u-blocks module based UNET and Inverted ReSidual Block cervix segmentation	DSC: 0.94
(Dash et al., 2023)	Colposcopy	292 images	Cervix segmentation	An inceptionNet with reduction feature with reduction mechanisms and area filtering for extracting cervix region	DSC: 0.8168
(Shinohara et al., 2023)	Colposcopy	30 subjects	Lesion segmentation	A UNET model for segmenting cervical lesion using acetic acid solution	AUC:0.943
(Yu et al., 2022a)	Colposcopy	5K images	Lesion segmentation	Multi-phased model with region of interest identification using Fast-RCNN to focus on main region of cervix, followed by lesion segmentation using EfficientNet-ASPP module.	DSC: 0.7371
(Li et al., 2023)	Colposcopy	1554 images	Segmentation and classification	A Deeplabv3+ module with Xception and ASPP module for classification and segmentation of cervical lesion.	IOU: 80.27
(Bai et al., 2018)	Colposcopy	100 subjects	Cervical region	A neighbouring average filtering and extracted hue, saturation, and value. Then k-means clustering for extracting cervical regions	IOU: 80.75
(Yuan et al., 2020)	Colposcopy	–	Lesion segmentation	A positive and negative classification using ResNet50, followed by UNET based segmentation and Masked R-CNN for lesion segmentation	DSC: 0.6164
(Bolaños Semanate et al., 2024)	Colposcopy	1050 images	Cervix region	Segmentation of cervix regions using Feature Pyramid Network	IOU: 0.8787
(Gao et al., 2025)	Colposcopy	9228 subjects	Lesion segmentation	A transformer-based network with boundary aware module and attention mechanism for Lesion segmentation	IOU: 0.6355
(Yang et al., 2025)	Colposcopy	971 subjects	Lesion segmentation	A Dense ASPP module with a multi-scale feature fusion for cervical lesion segmentation	IOU: 0.7340

Atrous spatial pyramid pooling (ASPP), Dice-Sørensen coefficient (DSC), intersection over union (IOU), region-based convolutional neural network (R-CNN), Feature Pyramid Network (FPN)

- Adaptively integrates multiscale contextual features, while balancing global context and subtle changes.
- Suppresses irrelevant background noise in skip connections to focus on cervix-relevant regions for lesion segmentation.
- Leverages a novel dataset with exact cervix mapping, ensuring reliable ground truth for the segmentation task.

Therefore, we propose a novel Cervix-Seg-Net model with a hybrid attention-based skip connection pyramid scene parsing (PSPNet) module, exploring pyramid pooling for global and regional context, attention mechanisms for adaptive weighting, and refined skip connections for noise reduction. We also introduce a unique dataset with accurate cervix mapping with precise ground truth for training and evaluation, providing a strong foundation for advancing the analysis of cervical lesions.

3. Methodology

3.1. Dataset

All data and images were collected from patients examined using a DYSIS colposcope Version 3 at the Department of Gynecology and Obstetrics, Randers Regional Hospital, Denmark, between 2017 and 2020. Colposcopy examinations were included if the colposcopists had reported a fully or partially visible transformation zone, and four cervical biopsies taken (Khare et al., 2025). Images not suitable for annotation were excluded (blurry, too much light, mucous, or blood covering the visible changes). The dataset comprised colposcopic images from 196 subjects who were referred for colposcopic examination after a screening test positive cytology or HPV test or as part of follow-up in women with previous dysplasia. The framework includes data from subjects who were identified as high-risk precancerous cervical lesions. In this category, four subclasses were included: cervical intraepithelial neoplasia (CIN2), CIN3, carcinoma in situ (CIS), and adenocarcinoma in situ (AIS). It should be noted that only one lesion type is identified per subject based on lesion severity. For a particular subject, if experts have identified two lesions, CIN2 and CIN3. The priority would be given to the most severe lesion, i.e., CIN3, while ignoring other lesions. This has resulted in a comprehensive dataset, containing various types of lesions learned by the proposed model.

3.2. Encoder Block

The first phase of the Cervix-Seg-Net model consisted of an encoder block, as shown in Fig. 1. We used a pre-trained EfficientNet-B3 model as the network encoder for the segmentation model. EfficientNet-B3 was selected due to its balance between accuracy and computational efficiency compared to other variants of EfficientNet, ResNet, and DenseNet. EfficientNet-B3 offers significantly fewer parameters and lower floating-point operations per second and achieves higher accuracy on the ImageNet dataset (Tan and Le, 2019). Furthermore, EfficientNet-B3 benefits from compound scaling by simultaneously increasing network

depth, width, and image resolution in a balanced manner and neural architecture search, enabling optimal scaling with enhanced performance for segmentation or classification (Yang et al., 2021; Tan and Le, 2019).

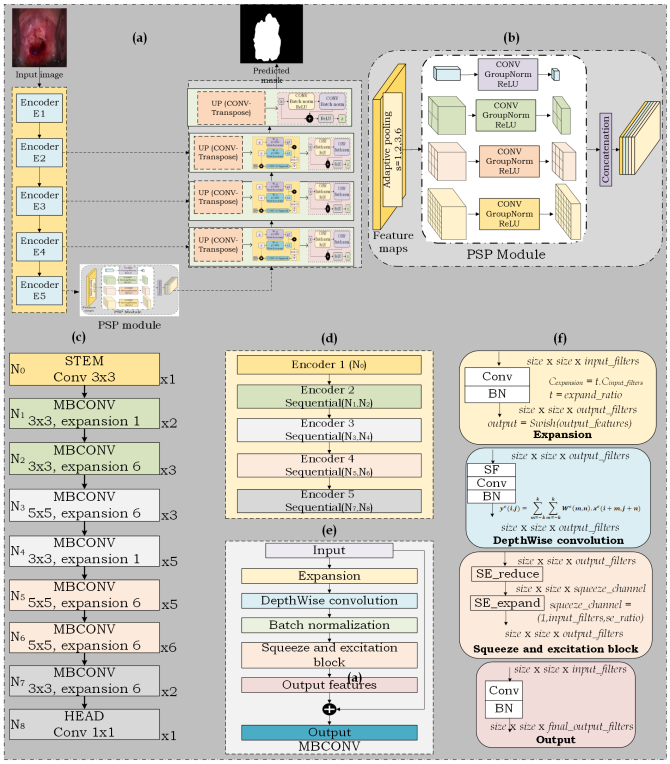

Figure 1: End-to-end pipeline of the Cervix-Seg-Net model. (a) schematic of the proposed hybrid Residual-Attention based skip-connection PSPNet; (b) schematic of PSP module; (c) architecture of EfficientNetB3; (d) the detail encoder block using MBCONV block; (e) each phase in a MBCONV block of EfficientNetB3; and (f) schematic diagram of MBCONV block.

The basic blocks of EfficientNet-B3 used for the encoder are composed of a stem, mobile inverted bottleneck convolution (MBCONV), and head modules, organized in nine nets as shown in Fig. 1 (c), which are composed of one stem, 26 MBCONV blocks, and one head. As indicated in Fig. 1 (d), these nets are organized in five encoders. The detailed schematic of each MBCONV block is shown in Fig. 1 (e), while the detailed explanation of each phase of the MBCONV block is shown in Fig. 1 (f). The proposed encoder block progressively reduces spatial resolution with increasing semantic abstraction. This allows for capturing local representations, like textures, shapes, and mid-level features. However, due to fixed convolutional kernels and pooling windows, these encoder features suffer from long-range dependencies and a missing global contextual prior.

3.3. PSP module

The Pyramid Scene Parsing (PSPNet) module captures the global contextual information across diverse scales within an image (Zhao et al., 2017). It uses the Pyramid Pooling Module (PPM) and adaptive pyramid scales as shown in Fig. 1. The coarsest (1×1) global pooling scale is used to generate a single bin output that captures the global representation in the image. The coarser intermediate scale is generated by a pooling of 2×2, while the finer scales with a pooling of 3×3 and 6×6 yielded more localized context representations. Thus, PPM output at various levels captures different contextual scales ranging from coarser to finer feature maps along with global receptive fields. Therefore, the final encoder features, which provide high-level semantic information, are passed through the PSP module to capture global and multi-scale contextual cues. This enriched representation provides the decoder with both scene-level understanding and localized details, guiding more accurate and consistent segmentation.

3.4. Decoder

The decoder block of the proposed Cervix-Seg-Net model combines three units: upsample, attention, and residual blocks, as shown in Fig. 1. The first 2D transposed convolutional unit in the decoder upsamples the incoming feature maps to improve the spatial resolution. The upsampled features are then passed through the attention block, which is explained in the following subsection.

3.4.1. Attention block

The proposed attention mechanism filters skipped features to suppress background noise and highlights relevant spatial regions in an input. First, the joint feature map with local and contextual information is obtained by:

$$\mathcal{H} = \mathrm{ReLU}(\mathcal{W}_g(g) + \mathcal{W}_x(x)) \tag{1}$$

where g and x are gating and skip features from the upsampling and encoder block. $\mathcal{W}_g$ and $\mathcal{W}_x$ are convolutional and batch normalization operations on decoder and encoder features, respectively. Following this, the attention-weighted feature maps are obtained by using attention coefficients, which are given by:

$$\begin{aligned} \alpha &= \sigma(\psi(\mathcal{H})) \\ x_{attention} &= x \odot \alpha \end{aligned} \tag{2}$$

where $\psi(\mathcal{H}) = \mathcal{W}_\psi(\mathcal{H})$ and $\mathcal{W}_\psi$ is a convolutional operation that reduces to a single channel, σ is a sigmoid activation, and $\odot$ is element-wise multiplication.

3.4.2. Residual block

Finally, the concatenated features obtained from upsampling and attention blocks are passed through a residual block for a detail-preserving reconstruction, facilitating gradient propagation and preventing degradation in deeper networks. The detailed formulation of the

residual block is represented by:

$$z = \mathrm{BN}_2\Big(\mathcal{W}_2 * \mathrm{ReLU}(\mathrm{BN}_1(\mathcal{W}_1 * x))\Big) \\ y = \mathrm{ReLU}(x + z) \tag{3}$$

where $*$ denotes convolution, BN is batch normalization, and y is a final feature map.

4. Experimental settings and results

The experimental settings comprised 3506 images from 196 subjects with high-risk lesions. A 64-bit operating system, 64 GB of RAM, an Intel(R) Xeon(R)-W5-2445 with a clock rate of 3.10 GHz, and the Nvidia RTX 4500 workstation with 24GB of GPU RAM were used to implement the model. To reduce the risk of overfitting and increase data variability during training, random geometric (flip, rotation, shift, scale, distortion) and color (hue-saturation, CLAHE, brightness/contrast) transformations have been applied. Training and validation have been performed using two strategies, namely holdout and five-fold cross-validation (FF-CV). The data during holdout validation was stratified into a 70:20:10 ratio for train:validation:test, yielding 2455, 700, and 351 images, respectively. For FF-CV, the entire dataset was divided into five parts; utilizing four parts for training and one for validation, which was iterated five times. We used 2806 images for training and validation of each fold, while 700 images were used for testing. The experimental setting was selected empirically with a batch size of 2, loss function as DiceBCE, a learning rate of $4{\times}10^{-4}$, number of epochs as 40, Adam optimizer, and a weight decay of $1{\times}10^{-6}$. It is noteworthy to mention that the experimental protocol was the same for each run and experiment.

We began evaluating our Cervix-Seg-Net model by comparing its Dice coefficient (DSC) with other benchmark segmentation models. We used UNET, UNET++, LinkNet, and PSPNet modules as baselines. For a fair comparison with the benchmark model, we used DenseNet121(DN121), ResNet24, and EfficientNet (EffNet)-B3 models. Table 2 summarizes the DSCs obtained on train, validation, and test sets along with tuning parameters for different models. As shown in Table 2, the Cervix-Seg-Net model yielded the highest DSC among other models in validation and test sets using holdout validation by three encoder variants. The EffNet-B3 variant provides the highest DSC of 0.811 and 0.748 for validation and test, respectively. The summary table shows that the EffNet-based UNET model requires fewer tuning parameters, having a total size of 159.34 MB. However, it provides the least DSC value for validation and test compared to other variants. On the other hand, the DN121-based variant of UNET++ requires the highest tuning parameters, with a total size of 606.96 MB, and provides DSC values for the training and test sets that are lower than those of the Cervix-Seg-Net model. These results highlight a balance between model complexity and performance. Models with few parameters exhibit limited DSC, whereas the model with the highest parameters offers marginal gains in DSC. The Cervix-Seg-Net model achieves superior DSC with 19.61M parameters (383.72 MB), demonstrating better generalization and practical efficiency for clinical applications.

To get insight into the model's performance, we also evaluated accuracy, intersection over union (IoU), precision, and recall for the Cervix-Seg-Net model using EfficientNet B3 as encoder. For the validation set, the model achieved an overall accuracy of 0.925,

Table 2: Comparison of DSC obtained during training, validation, and testing along with summary of tuning parameters for various models.

Model	Version	Train	Valid	Test	TP (M)	P (MB)	FP (MB)	TS (MB)
UNET	EffNet	0.821	0.784	0.697	6.25	23.2	135.36	159.34
	DN121	0.799	0.789	0.728	13.61	54.43	301.33	356.55
	ResNet34	0.811	0.778	0.740	24.44	97.75	143.65	242.19
UNET++	EffNet	0.818	0.778	0.710	13.62	51.79	273.48	326.06
	DN121	0.831	0.784	0.736	30.07	120.29	485.88	606.96
	ResNet34	0.798	0.790	0.720	26.08	104.31	265.29	370.39
LinkNet	EffNet	0.847	0.790	0.721	11.25	22.19	527.02	549.99
	DN121	0.784	0.782	0.728	19.26	77.05	257.43	335.26
	ResNet34	0.801	0.779	0.746	24.45	97.8	102.37	200.96
PSPNet	EffNet	0.813	0.756	0.737	18.34	73.38	276.08	350.24
Cervix-Seg-Net	ResNet34	0.808	0.790	0.742	26.86	107.44	428.72	536.95
	DN121	0.820	0.786	0.743	14.66	58.65	289.4	348.83
	EffNet	0.847	0.811	0.748	19.61	78.45	304.49	383.72

EfficientNet: EffNet; DenseNet121: DN121; TP: Total parameters; Parameters: P; Forward pass: FP; Total size: TS; million: M; megabyte: MB

IoU of 0.706, DSC of 0.811, precision of 0.796, and recall of 0.829, while for the test set, the model obtained an accuracy, IoU, DSC, precision, and recall of 0.896, 0.636, 0.748, 0.702, and 0.856, respectively. The precision and recall values demonstrated the balanced ability of Cervix-Seg-Net in identifying true positives while minimizing false detections. The gradual decrease in performance from training to test data shows a well-generalized model without significant overfitting, underscoring its robustness and applicability to unseen data. Finally, we validated our model's performance using 5-fold cross-validation to minimize bias and for better generalization. We reported the mean ± standard deviation (STD) of performance metrics with 95% confidence interval (CI), capturing central tendency and variability across folds, as shown in Table 3. As evident from the table, low STDs across all metrics during validation indicate stable performance. Also, t-stat and p-values confirm that improvements are unlikely due to random variation. Therefore, the results of Cervix-Seg-Net collectively indicate strong segmentation capabilities of the model in capturing subtle variations, providing an effective segmentation of high-risk cervical lesions.

5. Discussion

Fig. 2 represents the predictions generated by the Cervix-Seg-Net model. As shown in Fig. 2, each subfield indicates model input, GT, predictions yielded by the proposed model, and the error map. The figure reports resultant dice value (RDV), maximum pixel length, and pixel error, which presents a comprehensive understanding of each lesion. The figure shows that the Cervix-Seg-Net model is robust in segmenting larger lesions with well-defined boundaries, but its performance diminishes for smaller lesions. The model captures extra details around the cervix for smaller lesions in addition to fine structural details, resulting in an in-

Table 3: Results of five-fold cross validation obtained by Cervix-Seg-Net model.

	Train		Validation		Statistical test	
Metric	Mean ± STD	95%CI	Mean ± STD	95%CI	t-stat	p-value
Accuracy	0.95 ± 0.014	(0.933, 0.968)	0.904 ± 0.003	(0.9, 0.908)	6.984	0.002
IoU	0.813 ± 0.039	(0.765, 0.861)	0.673 ± 0.009	(0.662, 0.684)	7.997	0.001
DSC	0.89 ± 0.026	(0.858, 0.923)	0.783 ± 0.008	(0.773, 0.793)	8.938	0.001
Precision	0.875 ± 0.031	(0.837, 0.914)	0.773 ± 0.007	(0.764, 0.782)	7.038	0.002
Recall	0.909 ± 0.024	(0.879, 0.938)	0.795 ± 0.014	(0.777, 0.813)	12.562	0.002

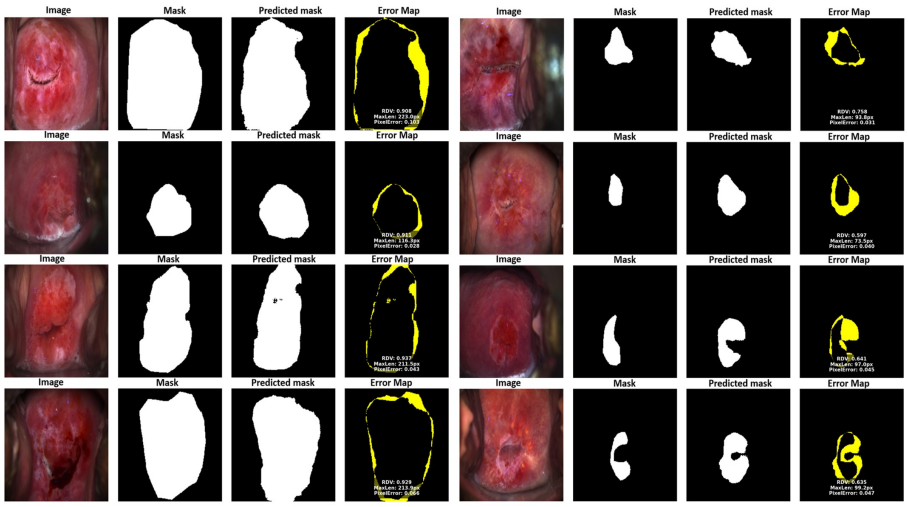

Figure 2: Examples of the predictions generated by the proposed Cervix-Seg-Net model. (a) Input image to the model; (b) Ground truth mask; (c) Predicted mask; and (d) Error map between ground truth and predicted mask.

herent trade-off between context aggregation and boundary precision. The model exhibits over-segmentation, producing additional boundaries that are not present in the ground truth, indicating an overestimation of structural details. In summary, our proposed model achieved superior performance compared to UNet, UNet++, LinkNet, and PSPNet. The EfficientNetB3 backbone in Cervix-Seg-Net facilitates hierarchical feature extraction, while the PSP module captures multi-scale contextual features, which are essential for identifying lesion boundaries. Additionally, the attention and residual blocks in the decoder enhance spatial representation and prevent feature degradation during upsampling. As a result, this hybrid architecture allows lesion localization, yielding superior segmentation performance and generalization compared to existing benchmark models. Our analysis indicates that high-quality, fine-grained GT annotations from clinical experts are still necessary to better guide the model's learning process for smaller lesions. A region-specific mapping within the cervix would enable the network to focus on the most diagnostically relevant features, like subtle epithelial changes or early lesion patterns, rather than generalizing across the entire cervical region. Such precise annotation would not only enhance the model's ability to capture clinically meaningful details but also improve its interpretability and reliability for targeted diagnostic assessment. Additionally, in the future, we aim to explore multi-scale attention mechanisms to enhance the segmentation of small, region-specific lesions, thereby improving technical generalization for targeted diagnostic assessment.

6. Conclusion

The paper presents a novel framework for segmenting high-risk cervical lesions. The proposed Cervix-Seg-Net model integrates multiscale context aggregation and attention-guided reconstruction, which yields effective lesion segmentation that captures clinically relevant features for high-risk cervical lesion segmentation. The performance comparison using hold-out validation reveals that the feature maps generated by our proposed model capture subtle variations within the cervix region more effectively than benchmark models, thereby yielding higher performance metric values. The lower standard deviation and statistically significant gains ($p < 0.05$) indicate the robustness and better generalization of the proposed model. The ability of Cervix-Seg-Net to effectively focus within the transformation zone allows robust delineation for downstream diagnostic and treatment planning. In conclusion, the proposed Cervix-Seg-Net framework potentially presents a technically efficient and clinically translatable solution for AI-assisted cervical lesion segmentation.

7. Ethical Statement

The study was approved by the Regional Health Research Ethics committee (journal number S-20190100), was registered with the Regional Data Protection Agency (journal number 19/29858) and with ClinicalTrials (identifier NCT04049357). All participants received verbal and written study information prior to participation and signed informed consent was obtained from each individual. The study was conducted in accordance with the declaration of Helsinki.

8. Acknowledgement

For this study we got funding from the Danmarks Frie Forskningsfond (DFF) (Grant number 4308-00058B).

References

Marc Arbyn, Elisabete Weiderpass, Laia Bruni, Silvia de Sanjosé, Mona Saraiya, Jacques Ferlay, and Freddie Bray. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. *The Lancet Global Health*, 8(2):e191–e203, 2020.

Bing Bai, Pei-Zhong Liu, Yong-Zhao Du, and Yan-Ming Luo. Automatic segmentation of cervical region in colposcopic images using k-means. *Australasian physical & engineering sciences in medicine*, 41(4):1077–1085, 2018.

Ana María Bolaños Semanate, Santiago Hurtado Bustos, Hernan Dario Vargas-Cardona, and Marcela Arrivillaga Quintero. Segmentation of the cervix in colposcopy images using machine learning techniques. In *2024 3rd International Congress of Biomedical Engineering and Bioengineering (CIIBBI)*, pages 1–5, 2024. doi: 10.1109/CIIBBI63846.2024.10784902.

Berit Bargum Booth, Lone Kjeld Petersen, Jan Blaakaer, Tonje Johansen, Henrik Mertz, Christina Blach Kristensen, Søren Lunde, Katja Dahl, and Pinar Bor. Dynamic spectral imaging colposcopy versus regular colposcopy in women referred with high-grade cytology: a nonrandomized prospective study. *Journal of lower genital tract disease*, 25(2):113–118, 2021.

Okpala Chibuike and Xiaopeng Yang. Convolutional neural network–vision transformer architecture with gated control mechanism and multi-scale fusion for enhanced pulmonary disease classification. *Diagnostics*, 14(24), 2024. ISSN 2075-4418. doi: 10.3390/diagnostics14242790.

Srikanta Dash, Prabira Kumar Sethy, and Santi Kumari Behera. Cervical transformation zone segmentation and classification based on improved inception-resnet-v2 using colposcopy images. *Cancer Informatics*, 22:11769351231161477, 2023. doi: 10.1177/11769351231161477.

Julia C Gage, Vivien W Hanson, Kim Abbey, Susan Dippery, Susi Gardner, Janet Kubota, Mark Schiffman, Diane Solomon, Jose Jeronimo, et al. Number of cervical biopsies and sensitivity of colposcopy. *Obstetrics & Gynecology*, 108(2):264–272, 2006.

Huayu Gao, Jing Li, Nanyan Shen, Wei Lu, Juanjuan Ma, and Ying Yang. Cervical lesion segmentation via transformer-based network with attention and boundary-aware modules. *Biomedical Signal Processing and Control*, 109:107946, 2025. ISSN 1746-8094. doi: https://doi.org/10.1016/j.bspc.2025.107946.

Yuxi He, Liping Liu, Jinliang Wang, Nannan Zhao, and Hangyu He. Colposcopic image segmentation based on feature refinement and attention. *IEEE Access*, 12:40856–40870, 2024. doi: 10.1109/ACCESS.2024.3378097.

Smith K. Khare, Berit Bargum Booth, Victoria Blanes-Vidal, Lone Kjeld Petersen, and Esmaeil S. Nadimi. An explainable attention model for cervical precancer risk classification using colposcopic images. *Computer Methods and Programs in Biomedicine*, 271:108976, 2025. ISSN 0169-2607. doi: https://doi.org/10.1016/j.cmpb.2025.108976.

Zhen Li, Chu-Mei Zeng, Yan-Gang Dong, Ying Cao, Li-Yao Yu, Hui-Ying Liu, Xun Tian, Rui Tian, Chao-Yue Zhong, Ting-Ting Zhao, Jia-Shuo Liu, Ye Chen, Li-Fang Li, Zhe-Ying Huang, Yu-Yan Wang, Zheng Hu, Jingjing Zhang, Jiu-Xing Liang, Ping Zhou, and Yi-Qin Lu. A segmentation model to detect cevical lesions based on machine learning of colposcopic images. *Heliyon*, 9(11):e21043, 2023. ISSN 2405-8440. doi: https://doi.org/10.1016/j.heliyon.2023.e21043.

Xuhang Lian, Yanwei Pang, Jungong Han, and Jing Pan. Cascaded hierarchical atrous spatial pyramid pooling module for semantic segmentation. *Pattern Recognition*, 110: 107622, 2021. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2020.107622.

Elayaraja P and Suganthi M. Automatic approach for cervical cancer detection and segmentation using neural network classifier. *Asian Pacific Journal of Cancer Prevention*, 19(12):3571–3580, 2018. doi: 10.31557/APJCP.2018.19.12.3571.

Jinhee Park, Hyunmo Yang, Hyun-Jin Roh, Woonggyu Jung, and Gil-Jin Jang. Encoder-weighted w-net for unsupervised segmentation of cervix region in colposcopy images. *Cancers*, 14(14), 2022. ISSN 2072-6694. doi: 10.3390/cancers14143400.

Toshihiro Shinohara, Kosuke Murakami, and Noriomi Matsumura. Diagnosis assistance in colposcopy by segmenting acetowhite epithelium using U-Net with images before and after acetic acid solution application. *Diagnostics*, 13(9), 2023. ISSN 2075-4418. doi: 10.3390/diagnostics13091596.

Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 71(3):209–249, 2021.

Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

Nicolas Wentzensen, Joan L Walker, Michael A Gold, Katie M Smith, Rosemary E Zuna, Cara Mathews, S Terence Dunn, Roy Zhang, Katherine Moxley, Erin Bishop, et al. Multiple biopsies and detection of cervical cancer precursors at colposcopy. *Journal of clinical oncology*, 33(1):83–89, 2015.

Barbara A Wuerthner and Maria Avila-Wallace. Cervical cancer: Screening, management, and prevention. *The Nurse Practitioner*, 41(9):18–23, 2016.

Yingxue Xu, Guihua Wen, Yang Hu, Mingnan Luo, Dan Dai, Yishan Zhuang, and Wendy Hall. Multiple attentional pyramid networks for chinese herbal recognition. *Pattern*

Recognition, 110:107558, 2021. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2020.107558.

Jiahui Yang, Ying Zhang, Wenlong Fan, Jie Wang, Xinhe Zhang, Chunhui Liu, Shuang Liu, and Linyan Xue. A novel lightweight multi-scale feature fusion segmentation algorithm for real-time cervical lesion screening. *Scientific Reports*, 15(1):6343, 2025.

Yuan Yang, Lin Zhang, Mingyu Du, Jingyu Bo, Haolei Liu, Lei Ren, Xiaohe Li, and M Jamal Deen. A comparative analysis of eleven neural networks architectures for small datasets of lung images of covid-19 patients toward improved clinical decisions. *Computers in Biology and Medicine*, 139:104887, 2021.

Hui Yu, Yinuo Fan, Huizhan Ma, Haifeng Zhang, Chengcheng Cao, Xuyao Yu, Jinglai Sun, Yuzhen Cao, and Yuzhen Liu. Segmentation of the cervical lesion region in colposcopic images based on deep learning. *Frontiers in Oncology*, Volume 12 - 2022, 2022a. ISSN 2234-943X. doi: 10.3389/fonc.2022.952847.

Hui Yu, Yinuo Fan, Huizhan Ma, Haifeng Zhang, Chengcheng Cao, Xuyao Yu, Jinglai Sun, Yuzhen Cao, and Yuzhen Liu. Segmentation of the cervical lesion region in colposcopic images based on deep learning. *Frontiers in Oncology*, Volume 12 - 2022, 2022b. ISSN 2234-943X. doi: 10.3389/fonc.2022.952847.

Chunnv Yuan, Yeli Yao, Bei Cheng, Yifan Cheng, Ying Li, Yang Li, Xuechen Liu, Xiaodong Cheng, Xing Xie, Jian Wu, et al. The application of deep learning based diagnostic system to cervical squamous intraepithelial lesions recognition in colposcopy images. *Scientific reports*, 10(1):11639, 2020.

Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6230–6239, 2017. doi: 10.1109/CVPR.2017.660.