Deep learning-based cervical lesion segmentation in colposcopic images

Lalasa Mukku [a,1,*], Jyothi Thomas [a,2]

[a] CHRIST(Deemed to be University), Kengeri campus, Bangalore – 560074, India
[1] mlalasa2020@gmail.com; [2] j.thomas@christuniversity.in
* corresponding author

ARTICLE INFO

Article history
Received March 09, 2024
Revised March 27, 2024
Accepted April 02, 2024
Available online April 03, 2024

Keywords
Segmentation
Cervical cancer
Colposcope
Artificial intelligence
Deep learning

ABSTRACT

Artificial intelligence assisted cancer detection has changed the ream of diagnosis precision. This study aims to propose a segmentation network using artificial intelligence for accurately segmenting the cervix region and acetowhite lesions in cervigram images, addressing the shortage of skilled colposcopists and streamlining the training process. A computational approach is employed to develop and train a deep learning model specifically tailored for cervix region and acetowhite lesion segmentation in cervigram images. A dataset acquired in collaboration with KIDWAI memorial cancer research institute is used for building the model. Cervigram images are collected for training and validation, and a deep learning architecture is constructed and trained using annotated datasets. The segmentation network based on efficientnet architecture and atrous spatial pyramid pooling is designed to accurately identify and delineate the target regions, with performance evaluation conducted using precision, accuracy, recall, dice score, and specificity metrics. The proposed segmentation network achieves a precision of 0.7387±0.1541, accuracy of 0.9291, recall of 0.7912±0.1439, dice score of 0.7431±0.1506, and specificity of 0.9589±0.0131, indicating its reliability and robustness in segmenting cervix regions and acetowhite lesions in cervigram images. This research demonstrates the feasibility and effectiveness of using artificial intelligence-based computational models for cervix region and acetowhite lesion segmentation in cervigram images. It provides a foundation for further investigations into classifying cervix malignancy using AI techniques, potentially enhancing early detection and treatment of cervical cancer while addressing the shortage of skilled professionals in the field.

1. Introduction

Cervical cancer stands as the second most prevalent cancer and a prominent cause of cancer-related mortality among women globally [1]. While developed nations exhibit a decreasing trend in cervical cancer incidence, developing and underdeveloped countries are witnessing a surge in the occurrence of cervical cancer [2], [3]. Colposcopy is a procedure during which a solution called acetic acid is smeared on the cervix region and a sequence of images of the cervix are captured using the probe [4]. These images are called cervigrams [5]. Cervigrams often contain surrounding organs like vaginal walls separated through a spatial dimension. In addition to that, there are undesired medical image equipment like speculum and stretcher. Precisely delineating critical organs or structures in medical images, including magnetic resonance (MR) [6], computed tomography (CT) [7], and cervigrams obtained through colposcopy, holds paramount importance in clinical practice. This significance is particularly

pronounced in the context of the escalating use of medical imaging for the diagnosis and treatment of various diseases, notably cancer [8], [9]. Segmentation precision not only paves the way for further quantitative evaluations of regions of interest it also proves instrumental in facilitating accurate diagnoses, prognostic predictions, and strategic planning for surgical interventions, including intra-operative guidance [10]. Presently, the acknowledged gold standard for segmentation outcomes is derived from the discerning expertise of seasoned physicians and radiologists who rely on visual scrutiny and manual delineation [11]. Due of its intrinsic subjectivity, manual labelling has a high potential for inter-observer variability and limited reproducibility. The segmentation quality outcomes are markedly impacted by the doctor's knowledge and experience in this regard [12]. Therefore, the imperative need for automated segmentation algorithms becomes evident, particularly when aiming to efficiently attain reproducible and precise results of segmentation in routine clinical practice. Automatic segmentation of cervix [13] is a formidably challenging task. Among patients, there is often notable variability in both the shape and appearance of the targeted objects [14]. To confront these challenges, a plethora of algorithms has been thoroughly investigated in the past decades. Level sets, active contours, multi-atlas, statistical shape modelling, and graphical models, with hand-crafted features—these have all been extensively explored to address the challenges posed by medical image segmentation. Yet, the representation capability of these handcrafted features often proves insufficient to address the extensive variations in appearance and shape. Subsequently, there has been an exploration of learning-based methods to harness more potent features. However, these methods still encounter challenges in fully leveraging the spatial information inherent in images to attain segmentation results of the desired quality. Natural image processing has seen a revolution recently thanks to convolutional neural networks (CNNs), which take advantage of their hierarchically obtained highly representative features [4], [5], [10]. Medical image analysis has benefitted significantly by this development [15]–[17]. Deep learning-methodologies emerged as a fierce and pivotal alternative arena to address conventional challenges in medical image segmentation tasks. Cervigram images exhibit significantly more intricate anatomical environments in comparison to CT scans and other modality medical images. Automated identification of cervical cancer from colposcope images using deep learning has garnered a significant amount of traction in the past two decades. The accuracy of diagnosis during colposcopy analysis is heavily reliant on the efficacy of cervix segmentation, making it a critical component in the training of deep learning models. Some models have attempted unsupervised segmentation with reasonable accuracy [10], [18]–[20]. Deep learning (DL) has displayed significant progress in image analysis tasks. Several priorly trained models that can be employed as backbone networks for image classification tasks have proven to be promising. Transfer learning is a system that uses DL models without having to develop and train them from scratch. A few DL models used to deal with a colposcope-based cervical cancer diagnosis are mask R-CNN, Faster R-CNN and Deeplab V3, Squeezenet, Resnet, VGG 16. The findings of these studies consistently report that the DL models have good performance accuracy at classification [21]. Nevertheless, several of those models work well on segmentation tasks as well. Machine learning (ML) has been a go-to solution for cancer diagnosis in the past couple of decades as well, and a considerable volume of research has been carried out to categorize abnormal lesions in the cervix [8]. Various ML-based image processing techniques like Gaussian mixture modelling, mean shift algorithm, canny edge detector, pyramid feature extraction, etc., proved to be efficient means to achieve the task. Table 1 displays the methods employed in cervix image segmentation over the past few years. Althogh the methods have performed satirsfactorily the support vector machine, mean shift algorithm, and canny edge detector have suffered from the methodological fallability of over segmentation. Although the K means and gaussian mixture models need further

research in terms of how to optimize hyperparameters like mixture co-efficient, eigen vector and covariance matrix.

Cervical cancer diagnosis presents a multifaceted challenge characterized by intricate anatomical structures, varied lesion appearances, and the inherent subjectivity of visual interpretation. The complexity of this disease is compounded by the limitations of conventional diagnostic methods, such as colposcopy and histopathology, which rely heavily on the expertise of clinicians and pathologists. Moreover, the detection and delineation of cervical lesions within medical imaging data, particularly in modalities like colposcopy images or cervical histology slides, demand precise segmentation to distinguish abnormal tissues from the surrounding normal structures accurately. The pivotal role of segmentation in this context lies in its capacity to facilitate the identification and characterization of subtle or ambiguous lesions, thereby enabling clinicians to make informed diagnostic and treatment decisions.

To overcome the said difficulties in cervix segmentation task, this paper presents a novel segmentation module that uses Faster RCNN in combination with a segmentation network. The main contributions and objectives of the paper are listed as: 1) In order to recover the cervical region and eliminate instrument and vaginal wall interference noise from the colposcope pose-acetic-acid images, a faster RCNN was employed; 2) It uses a novel cervical lesion segmentation network. The extraction of features from the cervical area was done using EfficientNet-B3; 3) The result of segmentation was conveniently transferred back to the original image for medical professionals to review. Following that, a heatmap was used to demonstrate the model's performance.

2. Method

The segmentation task is achieved in three steps. First, the cervix region is extracted. Subsequently a network for cervical segmentation is proposed. After the image has been restored, it is then mapped using the zoom ratio back to the original image. The proposed architecture is presented in Fig. 1.

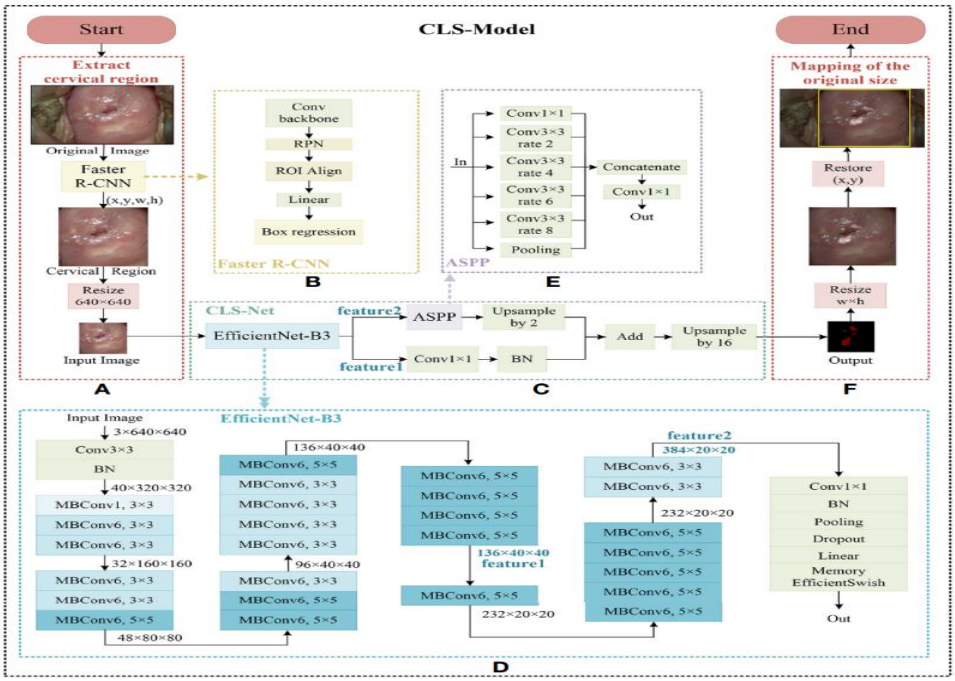

Fig. 1. Proposed architecture

2.1. Cervix Region of Interest Extraction

The cervical region was identified using the RCNN target detection method. In contrast to earlier segmentation techniques with uneven cervical borders and k-means clustering [22], the results of the segmentation performed in a rectangular format were beneficial for the simulations that followed. While the amount of time is similar, the enhanced Greater detection accuracy is provided by a faster R-CNN method that references Mask R-CNN and uses ROI Align technology than Faster R-CNN and other target detection methods. Due to the large size of the original colposcopic post-acetic-acid image and the extraction region, the extraction region was uniformly decreased to 640 × 640. The scale did not impair the lesion's later analysis. This made it easy to process the segmentation network that followed while upholding clarity. Fig. 1A depicts the general architecture of the cervical region extraction model, whereas Fig. 1B shows Faster R-CNN's enhanced structure. After the cervical area was removed, the coordinates of the rectangular frame's top left corner point, width, and height were documented; these coordinates are represented as (x, y, w, h).

The dataset contained 906 uterine cervix images of saline, acetic acid, and Lugol's iodine stages. The target classes are Normal, CIN 1 (cervical intraepithelial neoplasia), and CIN 2. Each case contained three cervix images captured by a colposcope in the time intervals 0 seconds, 60 seconds, and 120 seconds. In addition to that, clinical data pertaining to age, HPV test result, CIN grade, observations, proposed course of treatment etc., corresponding to images is available for each case. The data was split in an 80:20 ratio for training the model and testing it.

2.2. Cervix Segmentation Framework

The suggested model used an encoder-decoder structure that was in an end-to-end format; Fig. 1C depicts the general architecture. During the encoding phase, the cervical lesion region's features were extracted using the accurate and efficient EfficientNet-B3 [23] network. The proportions of the feature map were 1/32 of feature 2's original size on the 28th layer and 1/16 of feature 1's original size on the 20th layer. The two retrieved layers, which have superior deep pragmatic features, were the final layers below this size and in the subsampling procedure are two of the smallest sizes.

Low level characteristics included greater location and comprehensive information, along with a higher resolution. They contained more noise and less semantic information, nevertheless, because of the decreased convolution. Stronger semantic information was present in high-level characteristics, but they also exhibited poorer resolution and detail perception. Consequently, at the decoding phase, a multiscale feature fusion approach across layers was applied. After the convolutional layer and BN layer with a convolutional kernel of 1 × 1, the size of feature 1 was 1/16 of the original picture size. This was consistent with the 1/16 of the original size obtained by feature 2 with the ASPP module created for this lesion after two upsampling layers. After fusing the characteristic data from the nearby high and low layers, the sample was upsampled 16 times to its original size.

- The segmentation methodology utilized in this study employed an encoder-decoder structure, depicted in Fig. 1C, designed in an end-to-end format.
- Initially, during the encoding phase, features representing the cervical lesion region were extracted using the EfficientNet-B3 network, chosen for its balance of accuracy and efficiency.
- Specifically, the feature map proportions were set to 1/32 of the original size of feature 2 on the 28th layer and 1/16 of the original size of feature 1 on the 20th layer.

- These two layers, situated at the final stages of the encoding process and featuring superior deep pragmatic features, were identified as the most pertinent for segmentation due to their smaller size obtained through subsampling. Below is a detailed description of the ASPP module and the EfficientNet-B3 module, respectively.

2.2.1. EfficientNet Architecture

A standardized model extension technique called EfficientNet [24] achieves a great balance between the 3 parameters of model resolution, depth, and width. The squeezing method in SENet is utilized for network topology's optimization, and from MobileNetV2 [25], EfficientNet uses MBConv as the model's backbone network. The baseline model, EfficientNet-B0, was created using AutoML MNAS. In this work, feature extraction was accomplished using ImageNet pretrained EfficientNet-B3. Fig. 1D illustrates that of the 34 layers in the model, only the layer preceding the 28th layer was utilized in this work.

2.2.2. Atrous Spatial Pyramid Pooling

The ASPP [26] module uses a dilated convolution with several sampling speeds to sample the input feature graph in parallel. Then, using a 1 × 1 convolution, it concatenates the results to raise the channel number and decreases it to the output channels number. It also uses many scales to acquire visual feature information.

The ASPP module that was appropriate for the lesion's sample rate was redesigned in this paper using the DeepLabV3+ ASPP, which included six branches, one global average pooling four 3 × 3 dilated convolutions with rates{2,4,6,8}, and a 1 × 1 convolution,. This was done based on the lesion region's large area difference and size after subsampling. Next, in order to obtain multiscale information, we fused the features of six branches using a convolution fuse. Next, we lessened the channels to ½ of the original input layer using a 1 × 1 convolution. Fig. 1E displays the ASPP module's architecture.

2.3. Original image mapping

For the convenience of the colposcopists, the width and height coordinates (x, y, w, h) of the rectangular frame were used to map the segmentation image to the original image and upper-left corner point (y). This resulted in an image with a size of 640 × 640. Translucent black masks were used to overlay the normal region, while the lesion zone lacked any original appearance. As depicted in Fig. 1F.

3. Results and Discussion

This section presents and discusses the results obtained by the segmentation module proposed:

3.1. Dataset

The templ To assess the impact of our methodology for uterine cervix image segmentation, we utilized a dataset comprising 906 images representing saline smeared cervix, acetic acid smeared cervix, and Lugol's iodine smeared cervix stages. The dataset contains high resolution multistate cervix images taken with a colposcope of Field-of-View: 52° and Depth of View: 5mm - 120mm. Additionally, the clinical inferences are also present alongside the pathological report the result of which is used to label the classes. A Leisegang 3Ml LED colposcope with a Field-of-View of 52° and Depth of View ranging from 5mm to 120mm was employed. The target classes include Normal, CIN 1 (cervical intraepithelial neoplasia), and CIN 2. Each case in the dataset consists of three cervix images captured at time intervals

0 seconds, 60 seconds, and 120 seconds using a colposcope. Additionally, comprehensive clinical data, including information on age, HPV test results, CIN grade, observations, and proposed courses of treatment, is available for each case. The dataset was partitioned in an 80:20 ratio for testing and training the model.

3.2. Evaluation Metrics

The performance of several models was assessed using the five widely used criteria—dice, recall, accuracy, specificity and precision—which are as follows:

$$Dice = \frac{2 \times TP}{FP+FN+2\times TP} \tag{1}$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100 \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$Specificity = \frac{TN}{TN+FP} \tag{5}$$

$$Score = \frac{Dice+Recall}{2} \tag{6}$$

Here, FP denotes false positives FN denotes false negatives, TP stands for true positives and TN stands for true negatives. In the segmentation process, dice play a crucial role. Acetowhite lesions in cervix call for lower incidence of absented diagnoses nevertheless a reasonable percentage of misdiagnoses. As a result, the Specificity must be within a particular bracket and the Recall must be high.

3.3. Model Performance

Initially, we extracted the cervical region's features using the enhanced Faster R-CNN, and the AP@0.8 = 0.995 resulted. As a result, out of the 210 test sets, just one has an IOU lower than 0.8. Our conditions are satisfied by the AP@0.8 = 0.995. Fig. 2 displays the prediction box (green Pre) and ground truth box (red GT). In the original Faster R-CNN, AP@0.8 = 0.98. It is evident that applying ROI Align for correction results in an improvement in average precision (AP).

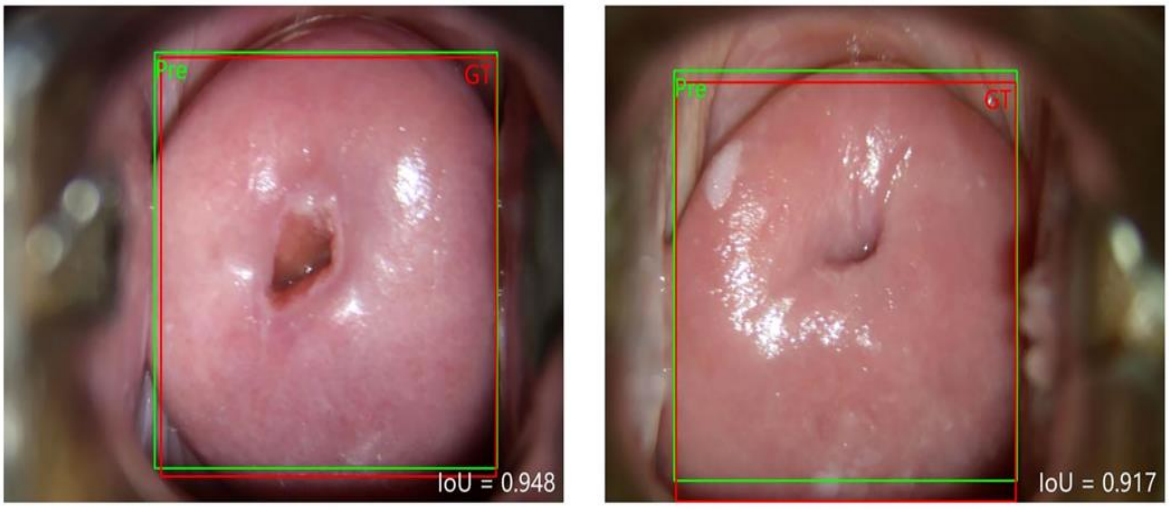

Fig. 2. Cervix region marking

Fig. 3 displays the validation and training loss alongside the score curves for the suggested model. At the 25th cycle, convergence begins between training and validation sets.

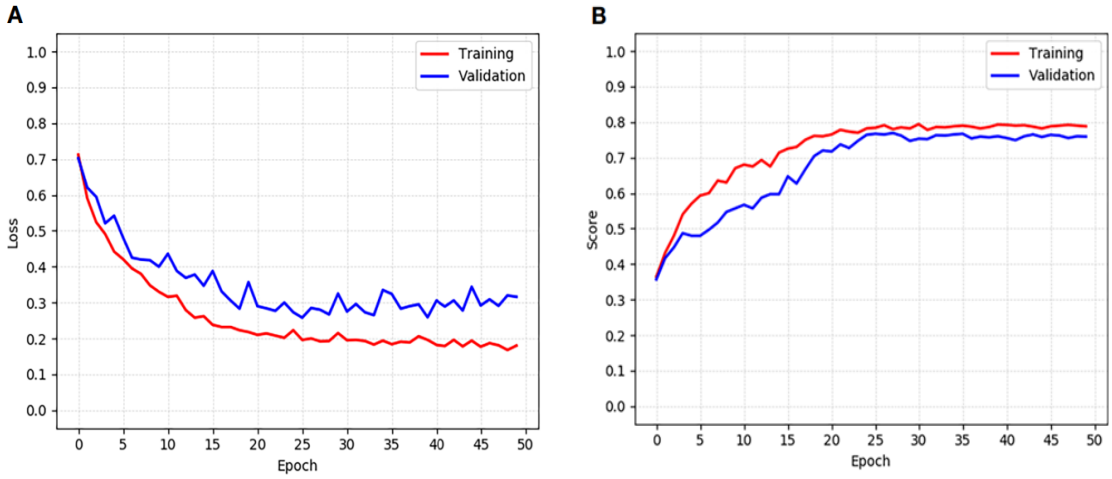

Fig. 3. A,B: Accuracy and loss training and validation curves

In the evaluation of segmentation performance, our proposed method was benchmarked against state-of-the-art techniques, including FCN8, UNet, SegNet, DeepLabV3+, and CCNet. Table 1 presents a comprehensive comparison of performance metrics such as Dice coefficient, Specificity, Recall, and others. Notably, our proposed model demonstrated superior performance across these metrics, outperforming existing methods. Particularly, the Dice coefficient and Accuracy are pivotal indicators in segmentation evaluation, with higher values signifying better segmentation quality. Our method achieved the highest Dice coefficient of 0.7431, surpassing UNet by 0.1124 and CCNet by 0.0582, consequently minimizing missed diagnoses. Additionally, all models exhibited high specificity values exceeding 0.95, indicating minimal misdiagnosis rates. Notably, our proposed framework achieved a precision of 0.7387, surpassing SegNet by 0.052 and CCNet by 0.0123, thus minimizing false positive cases and enhancing the reliability of positive predictions, thereby mitigating unnecessary therapeutic interventions for patients.

Table 1. Performance metrics with comparison with state of the art

Model	Precision	Accuracy	Recall	Dice	Specificity
DeepLabV3+ [27]	0.6889±0.2101	0.9083	0.6828±0.1945	0.6416±0.1816	0.9545±0.0167
CCNet[28]	0.7264±2.003	0.9191	0.7179±0.1898	0.6849±0.1802	0.9560±0.0196
FCN8x [29]	0.7102±0.2287	0.9094	0.6434±0.2097	0.6311±0.2059	0.9522±0.0185
UNet [30]	0.6941±0.2321	0.9073	0.6593±0.2233	0.6307±0.2175	0.9575±0.0223
SegNet [31]	0.6867±0.1898	0.9097	0.7057±0.1733	0.6600±0.1637	0.9517±0.0117
Proposed	0.7387±0.1541	0.9291	0.7912±0.1439	0.7431±0.1506	0.9589±0.0131

The results of the segmentation of the visible partial method are less accurate because, as Fig. 4 illustrates, they comprise a scaly normal SE (squamous epithelium) region that is extremely comparable to the acetowhite lesion zone. The segmentation findings of the approach suggested in have successfully distinguished between the lesion region and squamous epithelium region (Fig. 4).

The results of our study have several implications for existing literature in the field of cervical cancer diagnosis and segmentation. Firstly, our findings underscore the significance of leveraging advanced

segmentation techniques, such as the encoder-decoder architecture employed in our proposed methodology, in improving diagnostic accuracy. By achieving a Dice coefficient of 0.7431, our method surpasses previous state-of-the-art approaches, demonstrating its efficacy in accurately delineating cervical lesions from medical imaging data.

Furthermore, our study contributes novel insights into the performance of segmentation models, particularly in the context of cervical cancer diagnosis. While previous studies have reported on the effectiveness of various segmentation algorithms, our research extends this knowledge by highlighting the strengths of the proposed method in minimizing missed diagnoses and false positive cases. These findings suggest that our framework offers a more reliable and precise approach to cervical cancer segmentation, with potential implications for enhancing patient care and treatment outcomes.

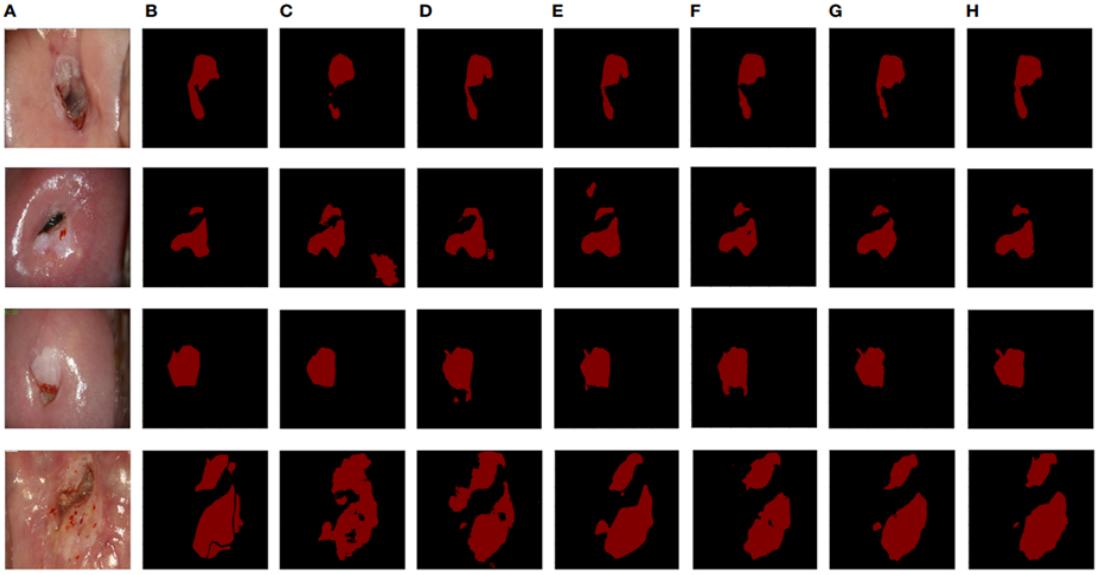

Fig. 4. Visualization A)Cervix image B) the ground truth C) Deeplab V3 D) CCNet E) FCN8X F) Unet G) SegNet H) Proposed method

4. Conclusion

In conclusion, this study addresses the pressing need for accurate cervical cancer diagnosis, particularly in regions with high disease burden. Our methodology focuses on delineating the cervix area from colposcope images, laying the groundwork for subsequent classification and automated cancer identification. Leveraging a Faster RCNN module for region extraction and a segmentation network for acetowhite segmentation, our approach demonstrates superior performance compared to existing methods. Through rigorous evaluation using standard metrics such as precision, Dice coefficient, accuracy, recall, and specificity, we validate the efficacy of our model. Importantly, our methodology's applicability to real-world clinical scenarios is highlighted by its ability to accurately identify cervical lesions, as confirmed against ground truth data. The presented framework holds significant promise for advancing uterine cervix image segmentation, with potential to enhance diagnostic accuracy and patient outcomes. Future research endeavors could explore integrating this framework as a preprocessing module for lesion classification, further augmenting its utility in clinical practice.

Declarations

Author contribution. LM conceptualized the study, implemented the code and prepared the manuscript. Prof JT supervised the experiment and guided the manuscript

Funding statement. None

Conflict of interest. The authors declare no conflict of interest.

Additional information. No additional information is available for this paper.

References

[1] A. D. Shrestha, D. Neupane, P. Vedsted, and P. Kallestrup, "Cervical Cancer Prevalence, Incidence and Mortality in Low and Middle Income Countries: A Systematic Review.," *Asian Pac. J. Cancer Prev.*, vol. 19, no. 2, pp. 319–324, Feb. 2018, doi: 10.22034/APJCP.2018.19.2.319.

[2] A. Ghoneim, G. Muhammad, and M. S. Hossain, "Cervical cancer classification using convolutional neural networks and extreme learning machines," *Futur. Gener. Comput. Syst.*, vol. 102, pp. 643–649, 2020, doi: 10.1016/j.future.2019.09.015.

[3] H. Sung *et al.*, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA. Cancer J. Clin.*, vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/caac.21660.

[4] S. Fang, J. Yang, M. Wang, C. Liu, and S. Liu, "An Improved Image Classification Method for Cervical Precancerous Lesions Based on ShuffleNet.," *Comput. Intell. Neurosci.*, vol. 2022, p. 9675628, 2022, doi: 10.1155/2022/9675628.

[5] Z. Yue *et al.*, "Automatic CIN Grades Prediction of Sequential Cervigram Image Using LSTM with Multistate CNN Features," *IEEE J. Biomed. Heal. Informatics*, vol. 24, no. 3, pp. 844–854, 2020, doi: 10.1109/JBHI.2019.2922682.

[6] R. Achalia *et al.*, "A proof of concept machine learning analysis using multimodal neuroimaging and neurocognitive measures as predictive biomarker in bipolar disorder," *Asian J. Psychiatr.*, vol. 50, p. 101984, 2020, doi: 10.1016/j.ajp.2020.101984.

[7] C. Yang, L. Qin, Y. Xie, and J. Liao, "Deep learning in CT image segmentation of cervical cancer: a systematic review and meta-analysis," *Radiat. Oncol.*, vol. 17, no. 1, p. 175, Nov. 2022, doi: 10.1186/s13014-022-02148-6.

[8] B. Hunter, S. Hindocha, and R. W. Lee, "The Role of Artificial Intelligence in Early Cancer Diagnosis," *Cancers (Basel).*, vol. 14, no. 6, p. 1524, Mar. 2022, doi: 10.3390/cancers14061524.

[9] X. Hou, G. Shen, L. Zhou, Y. Li, T. Wang, and X. Ma, "Artificial Intelligence in Cervical Cancer Screening and Diagnosis.," *Front. Oncol.*, vol. 12, p. 851367, 2022, doi: 10.3389/fonc.2022.851367.

[10] C. Yuan *et al.*, "The application of deep learning based diagnostic system to cervical squamous intraepithelial lesions recognition in colposcopy images," *Sci. Rep.*, pp. 1–12, 2020, doi: 10.1038/s41598-020-68252-3.

[11] M. Lalasa and J. Thomas, "A Review of Deep Learning Methods in Cervical Cancer Detection," in *International Conference on Soft Computing and Pattern Recognition*, 2022, pp. 624–633, doi: 10.1007/978-3-031-27524-1_60.

[12] A. Fourcade and R. H. Khonsari, "Deep learning in medical image analysis: A third eye for doctors," *J. Stomatol. oral Maxillofac. Surg.*, vol. 120, no. 4, pp. 279–288, 2019, doi: 10.1016/j.jormas.2019.06.002.

[13] L. Mukku and J. Thomas, "TelsNet: temporal lesion network embedding in a transformer model to detect cervical cancer through colposcope images," *Int. J. Adv. Intell. Informatics*, vol. 9, no. 3, pp. 502–523, 2023, doi: 10.26555/ijain.v9i3.1431.

[14] A. Traverso *et al.*, "Sensitivity of radiomic features to inter-observer variability and image pre-processing in Apparent Diffusion Coefficient (ADC) maps of cervix cancer patients," *Radiother. Oncol.*, vol. 143, pp. 88–94, Feb. 2020, doi: 10.1016/j.radonc.2019.08.008.

[15] S. Ali *et al.*, "A deep learning framework for quality assessment and restoration in video endoscopy," *Med. Image Anal.*, vol. 68, p. 101900, 2021, doi: 10.1016/j.media.2020.101900.

[16] K. A. Shastry and H. A. Sanjay, "Cancer diagnosis using artificial intelligence: a review," *Artif. Intell. Rev.*, vol. 55, no. 4, pp. 2641–2673, 2022, doi: 10.1007/s10462-021-10074-4.

[17] K. Sekaran, P. Chandana, N. M. Krishna, and S. Kadry, "Deep learning convolutional neural network (CNN) With Gaussian mixture model for predicting pancreatic cancer," *Multimed. Tools Appl.*, vol. 79, no. 15–16, pp. 10233–10247, Apr. 2020, doi: 10.1007/s11042-019-7419-5.

[18] H. Greenspan *et al.*, "Automatic detection of anatomical landmarks in uterine cervix images," *IEEE Trans. Med. Imaging*, vol. 28, no. 3, pp. 454–468, 2009, doi: 10.1109/TMI.2008.2007823.

[19] J. Liu, X. Sun, R. Li, and Y. Peng, "Recognition of Cervical Precancerous Lesions Based on Probability Distribution Feature Guidance.," *Curr. Med. imaging*, vol. 18, no. 11, pp. 1204–1213, 2022, doi: 10.2174/1573405618666220428104541.

[20] B. Bai, P.-Z. Liu, Y.-Z. Du, and Y.-M. Luo, "Automatic segmentation of cervical region in colposcopic images using K-means," *Australas. Phys. Eng. Sci. Med.*, vol. 41, no. 4, pp. 1077–1085, Dec. 2018, doi: 10.1007/s13246-018-0678-z.

[21] T. Zhang *et al.*, "Cervical precancerous lesions classification using pre-trained densely connected convolutional networks with colposcopy images," *Biomedical Signal Processing and Control*, vol. 55. pp. 1-11, 2020, doi: 10.1016/j.bspc.2019.101566.

[22] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci. (Ny).*, pp. 178-210, 2022, doi: 10.1016/j.ins.2022.11.139.

[23] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, 2019, pp. 6105–6114. [Online]. Available at: https://arxiv.org/pdf/1905.11946.pdf.

[24] M. Tan and Q. V Le, "EfficientNet: Improving Accuracy and Efficiency through AutoML and Model Scaling," *latest from Google Res.*, pp. 2–5, 2019. [Online]. Available at: https://blog.research.google/2019/05/efficientnet-improving-accuracy-and.html.

[25] C. Buiu, V.-R. Dănăilă, and C. N. Răduţă, "MobileNetV2 ensemble for cervical precancerous lesions classification," *Processes*, vol. 8, no. 5, p. 595, 2020, doi: 10.3390/pr8050595.

[26] Y. Qiu, Y. Liu, Y. Chen, J. Zhang, J. Zhu, and J. Xu, "A2SPPNet: attentive atrous spatial pyramid pooling network for salient object detection," *IEEE Trans. Multimed.*, pp. 1-16, 2022, doi: 10.1109/TMM.2022.3141933.

[27] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous Convolution for Semantic Image Segmentation." pp. 1-14, 2017. [Online]. Available at: https://arxiv.org/abs/1706.05587.

[28] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "Ccnet: Criss-cross attention for semantic segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 603–612, doi: 10.1109/ICCV.2019.00069.

[29] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440, doi: 10.1109/CVPR.2015.7298965.

[30] O. Ronneberger, P. Fischer, and 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, doi: 10.1007/978-3-319-24574-4_28.

[31] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, 2017, doi: 10.1109/TPAMI.2016.2644615.