

Title: Segmentation and Classification of Pap Smear Images to Detect Cervical Cancer using Deep Learning

Abstract

Cervical cancer remains one of the leading causes of death among women worldwide, and early detection is very important to reduce mortality. However, manual analysis of cervical cytology images is time-consuming, subjective, and highly dependent on expert pathologists, often leading to variability in diagnoses. Because of this limitation, automated computer-aided diagnosis (CAD) systems are needed to support more accurate and consistent screening. In recent years, deep learning (DL) has become an effective solution for medical image analysis because it can automatically learn complex visual patterns from data without manual feature extraction. In this study, DL-based methods are proposed for cervical cell segmentation and classification using three datasets, namely the private Pomeranian dataset, the publicly available SIPaKMeD dataset, and the Herlev dataset.

For segmentation, several DL architectures, including Improved U-Net, Spatial Pyramid Pooling (SPP)-SegNet, squeeze-and-excitation (SE-DeepLabV3+), Vision Transformer (ViT)-U-Net, and uncertainty-aware multi-tasking (UAMT) methods, are investigated. An Improved U-Net is used, which enhances the standard U-Net by refining feature extraction and improving boundary localization of cervical cells using dilation rates and Group Normalization. SPP-SegNet is applied to capture multi-scale contextual information for better segmentation of cells with varying sizes. SE-DeepLabV3+ introduces blocks to improve channel-wise feature recalibration and emphasize important feature responses. ViT-U-Net combines convolutional layers with ViT blocks to capture both local texture and global spatial relationships in cervical cell structures. Finally, UAMT is proposed, which performs segmentation and classification jointly while estimating prediction uncertainty to improve reliability in medical decision-making.

For classification, attention-based models, ensemble learning, and hybrid convolutional neural network (CNN)-ViT architectures are investigated to improve performance and robustness. An attention-based model is proposed, which adaptively focuses on the most relevant regions of cervical cells. A residual-based ensemble method combines multiple independently trained pretrained CNN models with residual refinement to improve fused feature representation and

prediction stability. A hybrid CNN-ViT model integrates CNN-based local feature extraction with transformer-based global dependency learning for robust classification. In addition, a learnable weight ensemble method automatically learns optimal weights for each model during training instead of using fixed averaging, improving overall performance. To address limited data availability, a residual deep convolutional generative adversarial network (RES_DCGAN) is used for data augmentation, generating realistic synthetic cervical cell images to increase dataset diversity and improve model robustness.

The proposed approaches are evaluated using train-test split, 5-fold cross-validation, and cross-dataset validation. The results show that the proposed methods consistently achieve high performance across most datasets. In most cases, segmentation improves interpretability by accurately localizing cell regions. In contrast, BBox-based inputs (bounding boxes with a small margin around segmented regions to include contextual information) perform better than ROI-based inputs (directly cropped segmented regions) by preserving useful surrounding context. The ensemble and attention-based models consistently outperform individual models, showing improved accuracy, stability, and generalization ability.

The proposed two-phase segmentation models with ensemble learning and transformer-based classification can significantly improve cervical cancer detection. In addition, the proposed model provides meaningful visual explanations and reliable decision support, making it suitable for real-world CAD. The results are competitive and, in several cases, outperform previous work, confirming the effectiveness of the proposed methods.

Keywords: Deep learning, Image Segmentation, Image Classification, Improved U-Net, Residual Ensemble, ViT, Data Augmentation, Cervical Cancer