Paper ID: 1570940294

Paper Title: Comparison of CNN- and Transformer-based Architectures for Automated Oral Epithelium Segmentation on Whole Slide Images

Authors: Napat Srisermphoak (Princess Srisavangavadhana College of Medicine, Chulabhorn Royal Academy, Thailand); Panomwat Amornphimoltham (Mahidol University, Thailand); Risa Chaisuparat (Chulalongkorn University, Thailand); Paniti Achararit and Todsaporn Fuangrod (Princess Srisavangavadhana College of Medicine, Chulabhorn Royal Academy, Thailand)

Email: napat.sri@edu.cra.ac.th

Abstract

Oral cancer is one of the most commonly found cancers worldwide. Oral Epithelial Dysplasia (OED) is an Oral Potentially Malignant Disorder (OPMD) that can be characterized for preventive oral cancer screening. The standard for OED histological grading is conducted via the epithelium regions of tissue biopsies. However, this procedure is laborious, time-consuming, and lacks reproducibility. Therefore, this study aims to explore the potential of using Convolutional Neural Network (CNN) and Transformer models for an automated epithelium segmentation algorithm directly from Whole Slide Images (WSIs).

This approach can reduce the manual process and support pathologists in grading activities. Accordingly, candidate architectures based on CNN and Transformer are selected: UNet, ResNet50-UNet, VGG19-UNet, Swin-UNet, and MISSFormer. These models are trained using patch-based segmentation to mitigate the high computational cost caused by processing WSIs. The results indicate that UNet, optimized with the ADAM optimizer, demonstrates the best performance in patch-based segmentation with an Intersection over Union (IoU) of 0.82 and a Dice-Similarity Coefficient (DSC) of 0.87. Furthermore, this model achieves the highest IoU and DSC for tissue-level prediction, scoring 0.88 and 0.94, respectively. After an in-depth analysis, the overlapping patching technique is considered to be disregardable for the majority of Neural Network architectures on tissue-level prediction. In return, this will lead to a substantial reduction in computational requirements.