An Integrated Approach to Automated Diagnosis of Cervical Intraepithelial Neoplasia in Digital Histology Images

Brid BROSnan, Inna SKARGA-BANDurova, Tetiana BILOBORODova and Illia SKARHA-BANDurov

a Oxford Brookes University, UK
b Saarland University of Applied Science HTW Saar, Saarbrücken, Germany
c University of Oxford, UK

Abstract. The study proposes an integrated approach to automated cervical intraepithelial neoplasia (CIN) diagnosis in epithelial patches extracted from digital histology images. The model ensemble, combined CNN classifier, and highest-performing fusion approach achieved an accuracy of 94.57%. This result demonstrates significant improvement over the state-of-the-art classifiers for cervical cancer histopathology images and promises further improvement in the automated diagnosis of CIN.

Keywords. Automated diagnosis, cervical intraepithelial neoplasia (CIN), histology image, deep learning, convolutional neural network (CNN), fusion

1. Introduction

Cervical cancer is a significant global challenge recorded as the fourth most common cancer among women globally, with estimated 604,000 new cases and 342,000 deaths in 2020 [1]. Cervical intraepithelial neoplasia (CIN) precedes invasive cervical cancer and begins with minimal structural abnormality progressing through stages of more significant abnormalities to invasive squamous cell carcinoma. Advances in AI and deep learning in digital pathology have been proven to have a significant improvement in diagnostic capabilities [2]. However, many AI-based solutions are still faced with many challenges. This project aimed to design an AI system able to cope with some of them and automate the detection of the CIN grade from digital histology images.

2. Methods

For this study, we used 1715 images from two datasets, own the CHI dataset [3] and samples obtained from the MTCHI dataset [4], to augment the CHI. MTCHI samples were derived from 80 digital images of histology whole slide images (WHI).
The technical context for automated CIN grading included the preparatory, detection, and decision-making phases. Data imbalance was handled, and samples were verified by expert pathologists. Seven ad-hoc CNN architectures were evaluated to define the best-performing architecture. Finally, the following data fusion techniques were applied to estimate the final CIN grade: A1 – the most severe CIN grade, even if detected in only a small part of the epithelium, is the confirmed stage of the disease; A2 – the most frequent occurrence of a grade; and A3 – an extension of A2 by using PCR rule 5 to determine predicted class confidence.

3. Results and Discussion

Repeated refinement of the data, and training parameters, based on evaluation results, enabled achievement of 92.14% accuracy for the model [5], further selected as a benchmark with mean F-score (92.19%), indicating excellent agreement between the model’s predictions and the true CIN values. Mean kappa (89.4%) and weighted kappa (90.83%) also indicate good agreement. These results suggest that the slight class imbalance is not a critical in this study, where CIN 2 is the majority class followed by CIN 1, Normal and CIN 3.

The cervical WSI image SSE extraction model [5] adopted for multiclass classification of histopathology images and combined with A2 fusion demonstrated the best performance to detect CIN 1 and CIN 2 in this dataset. However, the fusion method comparison was limited because only a few samples had conflicting predictions within their patch group. Individual misclassification findings include a non-neoplasm sample misclassified as CIN 3 by all fusion methods. One case was particularly extreme because the classifier predicts three different classes for its four patches, none of which are the actual class. Assessment of each CIN class shows that the CNN model best detects CIN 3. However, precision, the proportion of positive CIN 1 classes correctly identified, is 87.65%. Recall, the proportion of actual positive CIN 2 cases correctly identified, is 84.7%. Summing up the above, although overall performance is very good, the potential application of the classifier in medical diagnostics needs further investigation. Expert validation of the classifier output would provide more confidence in the model prior to fusing the patch results and assigning the fused predicted CIN grade to the parent sample. Ultimately, it is hoped that both stages of the complex histology image analysis project will be combined, enabling CIN grade identification directly from the histology sample image.

References