Multi-Task Deep Learning Model for Improved Histopathology Prediction from In-Vivo Microscopy Images

Anonymous Author(s)
Affiliation
Address
email

Cervical cancer remains a leading cause of death among women in low-to-middle income countries [4]. Breakthroughs in high-resource settings have improved diagnosis and treatment of cervical cancer, but these developments have not permeated into resource-constrained settings due to the high costs and the scarcity of trained healthcare professionals, colposcopists and pathologists [13]. Therefore, there is a need for low-cost devices that can be operated by nonspecialist healthcare providers to diagnose cervical cancer and its precursors at the point-of-care.

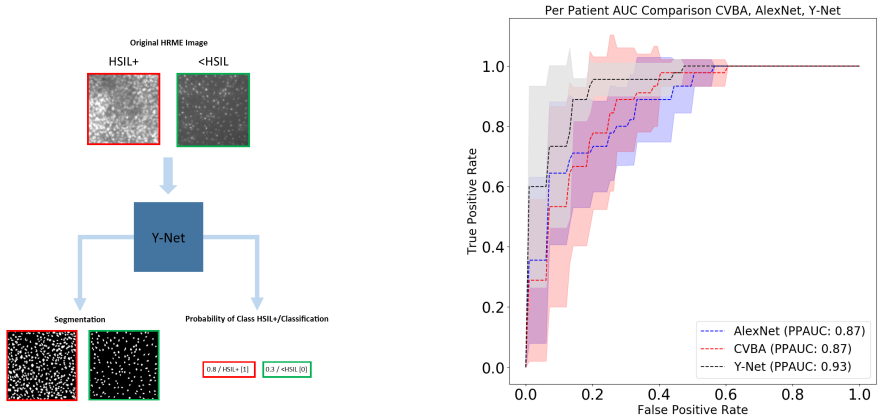
The High-Resolution Microendoscope (HRME) is a low-cost, in-vivo microscope that has been used for early detection of cervical cancer and precancer [10, 5]. Physicians apply proflavine dye on visually suspicious cervical lesions and use the HRME's optical probe to image nuclear morphology (Figure 1a). A computer-vision-based algorithm (CVBA) uses thresholding, watershed, and morphological operations to generate a nuclear binary mask. From this mask, abnormal nuclei are selected based on hand-crafted features such as nuclear eccentricity, nuclear size, and nuclear-to-cytoplasm ratio [11]. The number of "abnormal" nuclei per unit area is calculated for each patient image and serves as a diagnostic metric (HRME-score). The CVBA has shown results with sensitivity and specificity comparable to colposcopy when predicting pathology, but it depends on hand-crafted features that may not provide the best representation of the between-class distinctions in the dataset.

Convolutional Neural Networks (CNNs) have shown state-of-the-art performance by learning features for pathology prediction from stained slides and for label-free cytology analysis [1, 3]. These tasks involve data visually similar to that generated by the HRME, which suggests that the HRME system could also benefit from a deep learning approach. However, when working with medical imaging devices in a developmental phase such as the HRME, datasets are small, imbalanced, and noisy. Networks that excel in ImageNet [2] classification, such as AlexNet [8] and ResNet [6], perform on-par with the CVBA. AlexNet, the best performing model among the ResNet and AlexNet models tested, had a per patient area under the receiver operating characteristics curve (PPAUC) of 0.87 equal to that of the CVBA (Figure 1b). To improve on this performance, the HRME dataset requires network architectures that can compensate for data limitations by leveraging domain-specific knowledge in the form of auxiliary tasks for multi-task learning.

In this work, we present an exploration into how CNNs can be used to predict histopathology from HRME data. Experiments focus on demonstrating how using the Y-Net [9] architecture to learn the auxiliary task of nuclear segmentation along with histopathology classification improves prediction performance beyond that of standard CNN architectures such as AlexNet or ResNet.

Inspired by the CVBA's processing method, a network architecture was selected to mimic this system. Y-Net is a CNN architecture that performs simultaneous classification and segmentation. The architecture is based on U-Net [12] with an encoding path and a decoding path, joined by skip connections, that perform segmentation. However, Y-Net adds an additional branch emerging from the last encoding block in the segmentation branch to perform classification as shown in Figure 1a. This network is trained in two stages, first, the segmentation architecture is trained. Afterward, the diagnostic branch is added to the architecture and the network is trained for both classification and

Submitted to 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Do not distribute.



(a) Network architecture and sample images.

(b) PPAUC among the different algorithms with cross-validation standard deviation.

Figure 1: Y-Net architecture and performance on HRME histopathology prediction.

segmentation tasks. It is hypothesized that by leveraging features learned during the segmentation phase, the network's diagnostic performance will improve.

Implementing a Y-Net model for HRME-based pathology prediction required a segmentation mask target. Since it is difficult and resource-consuming to procure a reference segmentation for nuclear pixels, another method is proposed. In a self-supervised learning manner, Y-Net's segmentation branch was trained to generate the nuclear binary mask created by the CVBA. A high-level diagram of the network's architecture is shown in Figure 1a.

The HRME dataset was collected at Barretos Cancer Hospital in a rural area of Brazil [7]. 152 women underwent HRME in-vivo microscopy and cervical biopsies. After quality control, 124 patients (511 sites) were selected to form the dataset. Histopathology found 41 patients had inflammation, 38 had Cervical Intraepithelial Neoplasia (CIN) grade one (CIN1), 19 had CIN grade two (CIN2), 22 had CIN grade three (CIN3), and four had cancer. To act as ground truth, histopathology labels were binarized based on the treatment guidelines of the hospital such that sites and patients with inflammation or CIN1 pathology were referred to as less than high grade squamous intraepithelial lesion (<HSIL), while sites and patients with impressions of CIN2, CIN3, and cancer were regarded as greater than high grade squamous intraepithelial lesion (HSIL+).

Since the dataset is of limited size, k-fold cross-validation was performed where k is five. The dataset was randomly divided into five folds stratified by patient class labels. For each fold, the network is trained on four-fifths of the data and tested on one-fifth of the data. Only images with correlated site biopsy were used to train the network. Site biopsy pathology served as ground truth for training. When validating, all patient images are considered regardless of the existence of site biopsies. Patient-wise prediction is determined by selecting the worst image-wise diagnosis from a patient's set of HRME images. For the CVBA, patient prediction is the maximum HRME-score that the patient received. For the CNNs, patient prediction is the maximum HSIL+ class probability the patient received. The mean per patient receiver operating characteristics curves, with standard deviations for the cross-validation folds, are shown in Figure 1b for the best Y-Net model, the best AlexNet model, and the CVBA. PPAUC is used to compare Y-Net, AlexNet, and CVBA pathology prediction performance.

The highest PPAUC achieved by an AlexNet or ResNet model, including models pretrained and not pretrained on ImageNet, was 0.87 by an AlexNet model. This was equal to the CVBA mean PPAUC of 0.87. However, the Y-Net approach yielded a PPAUC of 0.93. This not only outperforms all previous CNN method used for the HRME dataset, but also surpasses the CVBA's performance. The CVBA uses an HRME-score threshold of 120, which corresponds to a sensitivity of 0.91 and a specificity of 0.67 for this dataset. At that same sensitivity of 0.91, the Y-Net model would provide a specificity of 0.81.

References

[1] Nicolas Coudray, Paolo Santiago Ocampo, Theodore Sakellaropoulos, Navneet Narula, Matija Snuder, David Fenyö, Andre L. Moreira, and Aristotelis Tsirigos Narges Razavian. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nature Medicine*, 24(10):1559–1567, 2018.

[2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255. Ieee, 2009.

[3] Minh Doan and Anne E. Carpente. Leveraging machine vision in cell-based diagnostics to do more with less. *Nature Materials*, 18(5):414–418, 2019.

[4] J Ferlay, M Ervik, F Lam, M Colombet, L Mery, M Pineros, A Znoar, I Soerjomataram, and F Bray. Cancer today.

[5] Benjamin Grant, José Fregnani, Júlio Possati Resende, Cristovam Scapulatempo-Neto, Graziela Matsushita, Edmundo Mauad, Timothy Quang, Mark Stoler, Philip Castle, Kathleen Schmeler, and Rebecca Richards-Kortum. High-resolution microendoscopy: a point-of-care diagnostic for cervical dysplasia in low-resource settings. *European Journal of Cancer Prevention*, 26(1):63–70, 2017.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[7] Brady Hunt, José Humberto Tavares Guerreiro Fregnani, Richard A Schwarz, Naitielle Pantano, Suelen Tesoni, Júlio César Possati-Resende, Marcio Antoniazzi, Bruno de Oliveira Fonseca, Graziela de Macêdo Matsushita, Cristovam Scapulatempo-Neto, et al. Diagnosing cervical neoplasia in rural brazil using a mobile van equipped with in vivo microscopy: A cluster-randomized community trial. *Cancer Prevention Research*, 11(6):359–370, 2018.

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 1097–1105, 2012.

[9] Sachin Mehta, Ezgi Mercan, Jamen Bartlett, Donald Weaver, Joann G Elmore, and Linda Shapiro. Y-net: Joint segmentation and classification for diagnosis of breast biopsy images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 893–901. Springer, 2018.

[10] Mark C Pierce, YaoYao Guan, Mary Kate Quinn, Xun Zhang, Wen-Hua Zhang, You-Lin Qiao, Philip Castle, and Rebecca Richards-Kortum. A pilot study of low-cost, high-resolution microendoscopy as a tool for identifying women with cervical precancer. *Cancer Prevention Research*, 5(11):1273–1279, 2012.

[11] Timothy Quang, Richard A. Schwarz, Sanford M. Dawsey, Mimi C. Tan, Kalpesh Patel, Xinying Yu, and Guiqi Wang. A tablet-interfaced high-resolution microendoscope with automated image interpretation for real-time evaluation of esophageal squamous cell neoplasia. *Gastrointestinal endoscopy*, 84(5):834–841, 2016.

[12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241. Springer, 2015.

[13] Michael L Wilson, Kenneth A Fleming, Modupe A Kuti, Lai Meng Looi, Nestor Lago, and Kun Ru. Access to pathology and laboratory medicine services: a crucial gap. *The Lancet*, 391(10133):1927–1938, 2018.