Classification of Epithelial Ovarian Carcinoma Whole-Slide Pathology Images Using Deep Transfer Learning

Abstract
Ovarian cancer is the most lethal cancer of the female reproductive organs. There are 5 major histological subtypes of epithelial ovarian cancer, each with distinct morphological, genetic, and clinical features. Currently, these subtypes are determined by a pathologist’s microscopic examination of tumor whole-slide images (WSI). This process is hampered by poor inter-observer agreement (Cohen’s kappa 0.54-0.67). We developed a two-stage deep transfer learning algorithm based on convolutional neural networks (CNN) for automatic classification of epithelial ovarian carcinoma WSIs. The proposed algorithm achieved a mean accuracy of 87.54% and Cohen’s kappa of 0.8106 in the slide-level classification of 305 WSIs; performing better than a standard CNN and pathologists without gynecology-specific training.

Keywords: Transfer learning, Ovarian cancer, Digital pathology

1. Introduction
Ovarian cancer ranks fifth in cancer deaths among women (Siegel et al., 2016), accounting for more deaths than any other cancer of the female reproductive system in North America. There are 5 major histological subtypes of epithelial ovarian cancer: high-grade serous ovarian carcinoma (HGSOC), clear cell ovarian carcinoma (CCOC), endometrioid (ENOC), low-grade serous (LGSOC) and mucinous carcinoma (MUC). These five major histotypes have distinct morphological, molecular, genetic, and clinical features (Köbel et al., 2008). Accuracy of pathologists in ovarian cancer histotype classification based on histo-morphological features of the tissue is hampered by poor diagnostic reproducibility and interobserver disagreement (Gilks et al., 2013; Clarke and Gilks, 2010; Han et al., 2013). Without gynecology-specific training, which reflects most current pathology practices, the interobserver agreement is moderate, with Cohen’s kappa varying between 0.54 and 0.67 (Köbel et al., 2013; Patel et al., 2012). Deep learning applied to image analysis, enabled by the digitization of pathology materials, provides an opportunity to revisit the rich information present in histopathology images, and improve ovarian cancer diagnosis.

2. Dataset
We collected and annotated a dataset of 305 whole-slide images (WSI) composed of 157 HGSOC, 53 CCOC, 55 ENOC, 29 LGSOC, and 11 MUC slides. The aforementioned WSIs originated from 159 ovarian cancer patients (76 HGSOC, 32 CCOC, 28 ENOC, 14 LGSOC,
Representative areas of tumor in each WSI were annotated by a board-certified pathologist. Due to the prohibitively large size of the WSIs, we tiled these annotated regions to patches of size 1024 × 1024 pixels at 40× magnification and down-sampled them to 512 × 512 and 256 × 256 using Lanczos filter. For evaluation, we utilized a 3-fold cross-validation scheme, in which we randomly divided the dataset into three patient groups. Two of the three groups were used as the training set, with the remaining group divided equally by patient and alternatively swapped for validation and test sets.

3. Method and Results

Most classification methods for WSI employ a patch-based approach in which the classification is performed on every single patch (e.g., 256 × 256 pixels) and an aggregation method (e.g., majority vote) is used to predict the label of the WSI from all the patch-level labels. However, a small patch with a limited field-of-view (FOV) might lose the context of the morphological patterns which are required to accurately classify a WSI. To extend the FOV while considering the computational limits, tiled patches are usually down-sampled from high-resolution to low-resolution. The trade-off between the size of FOV and the resolution makes it difficult to balance the context and details in images. Hence, this paper proposes a two-stage deep transfer learning patch-level classification method (Figure 1), which is not only able to extract the features from large FOV low-resolution patches, but is also able to further extract the small details from high-resolution patches. Finally, we aggregated the patch-level classification to slide-level classification using ensemble learning.

In Stage 1, we fed the low-resolution patches (i.e., 256 × 256 pixels) into a pre-trained VGG19 (Simonyan and Zisserman, 2015) network and replaced the last 1000-class fully-connected (FC) layer with a 5-class FC layer. A softmax layer was then applied at the end to obtain the categorical distribution corresponding to the five subtypes of ovarian cancer. Although a network trained on low-resolution images will potentially miss subtle features available from high-resolution images, it will likely pick up high-level contextual patterns that are critical for diagnosis and only available in larger FOV.

In Stage 2, we removed the first convolutional block of VGG19 and added two randomly initialized convolutional blocks at top of the trained network from Stage 1 and fed the high-resolution patches (i.e., 512 × 512 pixels) into the resulting network. Stage 2 takes advantage of the contextual features learned from low-resolution images, plus other essential small details from high-resolution images to further increase performance.

To aggregate the patch-level classification to slide-level classification, we created a matrix $Z^{N \times C}$, where $N$ is the number of WSIs and $C$ is the number of subtypes. We assigned each patch the label corresponding to the results of the Stage 2 classifier. Finally, to predict the WSI-level labels, we employed a random forest classifier.

To demonstrate the utility of our proposed approach, we compared the performance to a conventional CNN in which we trained a VGG19 network using 512 × 512 patches. Table 1 shows the per-class and overall performance of the patch-level and slide-level classifiers.
Transfer Learning-Based Histopathology Image Classification

4. Conclusion

We proposed a transfer learning approach for the automatic histological classification of epithelial ovarian cancers. The proposed algorithm performed better than standard CNNs (across various performance measures) as well as pathologists without gynecology-specific training. These results suggest a promising future direction to validate the findings on a larger cohort of patients. Code available at (link withheld).

References


