

UBC-OCEAN

hsharma@bu.edu, sbit@bu.edu

October 2023

Basic Information

Team members

- Harsh Sharma, hsharma@bu.edu
- Subhrangshu Bit, sbit@bu.edu

1 Introduction

Ovarian carcinoma is the most lethal cancer of the female reproductive system. There are five common types of ovarian cancer -

- High-grade serous carcinoma (HGSC)
- Clear-cell ovarian carcinoma (CC)
- Endometrioid (EC)
- Low-grade serous carcinoma (LGSC)
- Mucinous carcinoma (MC)

Additionally, there are several rare subtypes (out-of-distribution - "Outliers"). These are all characterized by *distinct cellular morphologies, etiologies, molecular and genetic profiles, and clinical attributes*.

Subtype-specific treatment approaches are gaining prominence, though first requires subtype identification, a process that could be improved with data science. Currently, ovarian cancer diagnosis relies on pathologists to assess subtypes. However, this presents several challenges, including disagreements between observers and the reproducibility of diagnostics. Furthermore, underserved communities often lack access to specialist pathologists, and even well-developed communities face a shortage of pathologists with expertise in gynecologic malignancies. Therefore, making the problem of accurate assessment of the subtypes an important addition to pathologists.

2 Data Analysis

The data was collected from 20 centers across 4 continents. The data set comprises two main categories of images:

- Whole Slide Images(WSI): 20x magnification and quite large.
- Tissue Microarray(TMA): 40x magnification and roughly 4000×4000 .

However, for this particular study, these images have been downsampled and converted to the Joint Photographic Experts Group(JPEG) format for easier computation and developing models on simpler computer configurations. The test set contains images from different sources hospitals than the train set, with the largest area images almost $100,000 \times 50,000$.

Challenges and considerations:

1. **Dimension Variability:** Variations in image dimensions necessitate a standardized representation of fixed shape for analysis.
2. **Quality and Staining Techniques:** Variances in image quality and slide staining techniques introduce potential variability.

The dataset exhibits significant class imbalance, as depicted in Table 1, showcasing varied counts against each label category.

Label	Count
HGSC	222
EC	124
CC	99
LGSC	47
MC	46

Table 1: Distribution of image counts

3 Methods

3.1 Data Preprocessing

Data preprocessing is essential to segregate background regions that do not contribute to the model's outcomes but consume computational resources.

3.1.1 Utilization of OpenCV Bradski [2000] Library

For data preprocessing, we employed the widely used OpenCV library, leveraging its existing functions. The steps undertaken in our preprocessing pipeline are illustrated in Figure 1:

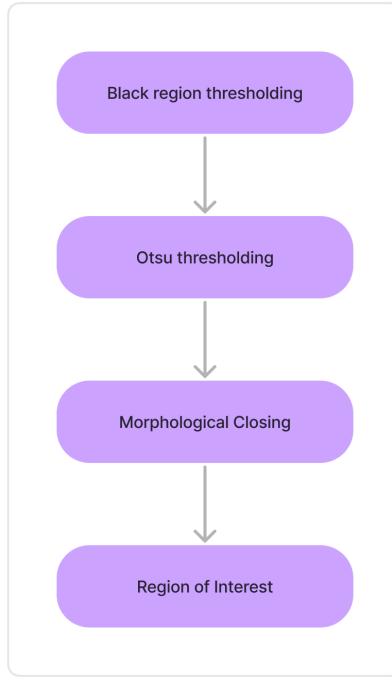


Figure 1: Preprocessing pipeline used

1. **Thresholding Black Regions:** Removal of black regions in the images devoid of pertinent information.
2. **Otsu Thresholding:** This adaptive thresholding (Otsu [1979]) method identifies potential information-rich regions based on image intensity. By minimizing intra-class intensity variance or maximizing inter-class variance, this technique determines regions of interest. Figure 2 provides an illustrative example of its application on a sample image.
3. **Morphological Closing:** To ensure cohesive structures representing distinct tissues, a closing operation is applied. This step efficiently addresses any discrepancies in region identification by the Otsu method.
4. **Identification of Regions of Interest:** Post-closure, a minimum size criterion is applied to the identified closed regions and holes. This criterion aims to filter out potential noise mistakenly identified by the Otsu method as tissue.

Figure 3 visually demonstrates the resultant tissue representation after the completion of our preprocessing pipeline.

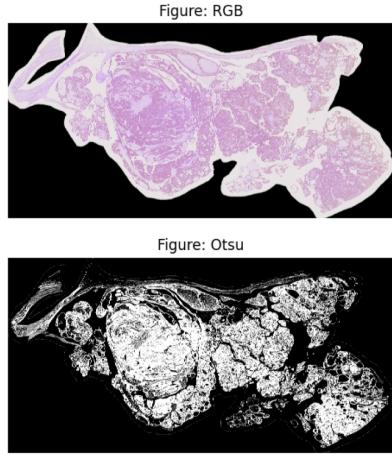


Figure 2: Example of OTSU thresholding on a RGB image

3.1.2 Patching

Our subsequent step involved the implementation of patching techniques, driven by insights from Zheng et al. [2022] and Lu et al. [2021], which highlight the challenges of managing whole slide images in contemporary research. Due to their unwieldy sizes, patching has emerged as a practical solution, mirroring methodologies employed in cutting-edge approaches for handling such images.

We proceeded with a uniform patching approach, employing a fixed patch size of 512 x 512 to segment the image for further analysis. Shown in Figure 4 is an example of how patching would be done.

3.2 Feature Extraction

Following our patch generation phase, the subsequent crucial step involved feature extraction from the generated patches. To achieve this, we surveyed a range of state-of-the-art models widely employed for feature extraction, particularly focusing on their utilization within the domain of whole slide images.

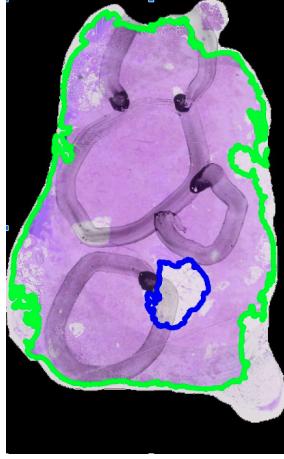


Figure 3: Example image with foreground-background separation, green shows region to be included, blue shows holes in the images which would not be processed

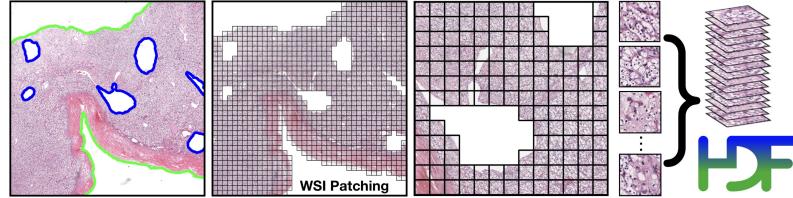


Figure 4: Example of image patches created from the foreground selected, Image reference: Lu et al. [2021]

3.2.1 Swin transformer-based pretrained model

The CTransPath model Wang et al. [2022], a hybrid architecture combining a convolutional neural network (CNN) and a multi-scale Swin Transformer, was introduced as a self-supervised learning (SSL) strategy for histopathological image analysis. It leverages semantically relevant contrastive learning (SRCL) to extract informative representations from unlabeled data, addressing challenges posed by limited annotations in medical image analysis.

Key highlights of the CTransPath model include:

1. **Semantically-Relevant Contrastive Learning (SRCL):** Enhances SSL by aligning multiple positive instances sharing similar visual concepts, fostering more diverse and informative representations.

2. **Hybrid Architecture:** Integrates a CNN and a multi-scale Swin Transformer, enabling stable network training and facilitating a powerful feature extractor with fine local structure and global context.
3. **Unsupervised Pretraining:** Utilizes massively unlabeled histopathological images for unsupervised pretraining, establishing a collaborative local-global feature extractor suitable for histopathology tasks.
4. **Performance and Transferability:** Demonstrates state-of-the-art performance in various downstream tasks across nine public histopathological datasets, exhibiting robustness and superior transferability compared to other SSL methods and ImageNet pretraining.

For our study, we aim to utilize the pretrained CTransPath model as a feature extractor for patch-level analysis. Its ability to derive universal and informative representations from histopathological images aligns with our objective of extracting meaningful features from our dataset.

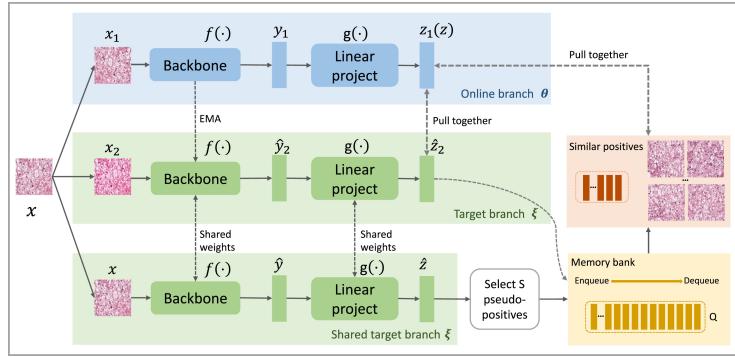


Figure 5: Contrastive learning method used in cTranspath, Image from Wang et al. [2022]

3.3 Residual Network

Residual Network (ResNet) He et al. [2015] is a convolutional neural network that introduces the concepts of residual learning and skip connections which enables training much deeper models without the problem of vanishing gradients. The model showed state-of-the-art performance in multiple benchmark datasets in 2015. There are multiple variants of the model architecture with varying numbers of layers. In this study, we leverage a ResNet50 model which is pretrained on the ImageNet dataset. However, based on Lu et al. [2021] the model architecture is slightly customized by dropping the third residual block and using adaptive mean-spatial pooling to convert each of the preprocessed 512×512 patches into 1024-dimensional feature vectors.

3.4 Classification models

3.4.1 CLAM

Based on one of the recent standard and robust frameworks CLAM developed by Lu et al. [2021], we leveraged similar methods to our problem. The extracted features from multiple patches of a whole slide image form a bag of multiple instances, thereby motivating the usage of multi-instance learning (MIL). However, since the data does not have the information on the labels of every patch, general MIL settings assume extrapolating the slide level label to each patch or using some sort of pooling functions such as max or majority to come up with a single label. These operators are of limited flexibility, non-trainable, and require problem-specific tuning. Thus, this approach leverages the attention-based MIL from Ilse et al. [2018] to aggregate slide-level representations from patch-level representations for each class. The framework allows single-branch attention for binary classification or multiple-branch attention heads, each head corresponding to one class for multi-class settings. This enables the network to learn from each class, allowing it to clearly distinguish between positive and negative evidence for each class.

To further constrain the patch-level feature space, the approach incorporates an additional binary clustering objective during training. For each of the n classes, since the patch-level labels are unknown, pseudo labels are generated based on the top best and worst attention scores, which are expected to represent the strong and weak evidence respectively. The objective is chosen to be the smooth Support Vector Machine (SVM) loss which allows a margin of error over strict cross-entropy loss.

3.4.2 Graph Based methods

Inspired by the similar work detailed in Zheng et al. [2022], which underscores the efficacy of constructing graph representations from images for superior performance while circumventing high computational requirements, we embarked on a path to harness similar principles.

Our objective involves the creation of graph structures derived from patches within the regions of interest, following methodologies akin to those elucidated in the referenced paper. By employing these methods, we aim to generate feature representations from these graph structures, capitalizing on the insights gleaned from the innovative approaches detailed in the aforementioned study.

This approach allows us to exploit the inherent information encapsulated within the interconnectedness of patches, facilitating a more comprehensive understanding of the image at both local and global levels. By integrating these graph-based representations with various graph neural network models, we anticipate augmenting our capacity to extract meaningful features and enhance the analysis of pathology images, all while optimizing computational resources.

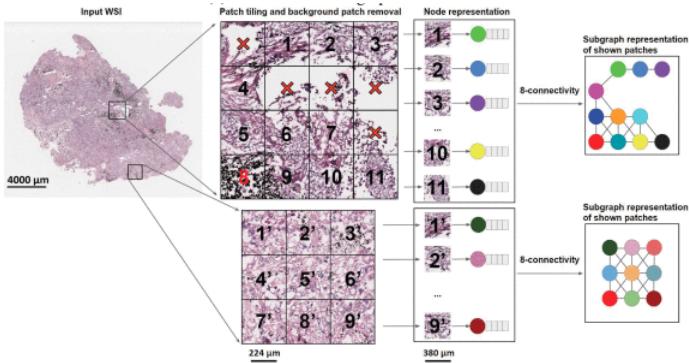


Figure 6: Graph Transformer Zheng et al. [2022]

4 Experiments and Results

Submissions are evaluated using **balanced accuracy**.

The table below presents the highest accuracies achieved during our experimentation with architectures inspired by the methodologies referenced earlier. We explored frameworks derived from the aforementioned methods to derive insights into their performance.

Prediction model	Feature Extraction model	Balanced Accuracy, mean% (std%)
CLAM	Resnet50	65.2 (5.2)
CLAM	cTranspath	74.09 (7.3)
Graph Transformer	Resnet18	68.5 (3.8)
Graph Transformer	cTranspath	77.3 (5.7)

Table 2: Performance evaluation

5 Discussion

Achieving a test score of 73% on the hidden test data set for this competition marks a notable milestone; however, significant strides remain imperative within this domain, primarily concerning computational demands. The development of more compute-efficient models is pivotal for enhancing the efficacy of these models in assisting pathologists and streamlining their workflow.

Our observations unveiled substantial intra-model variations, likely stemming from diverse sources. Images sourced from various machines and laboratories, featuring distinct staining methods, tissue slice thicknesses, and image acquisition devices, contribute to a notable lack of model generalizability.

Addressing these challenges—pertaining to compute efficiency and accommodating diverse image origins—stands pivotal for advancing the applicability

and robustness of models in clinical settings.

Moreover, during our investigation, we noted unstable training of the graph neural network. To mitigate computational demands, we adopted a strategy of selecting subsections of the image during each epoch for training the graph transformer. While intended to improve efficiency and encourage generalization based on subsections, this approach might be inadequate for localized cancer types. We suspect that the instability in training the network could be attributed to the absence of cancerous regions in some selected subsections, hampering its ability to learn effectively from diverse cancer presentations. This limitation warrants further exploration and refinement in training methodologies to address localized variations within histopathological images.

6 Conclusion

Our final submission strategy involved aggregating model probabilities across each fold. The model selected for submission was determined by averaging these probabilities, ultimately choosing the model with the highest probability. This approach aimed to leverage ensemble learning for a more robust and reliable final model selection. We used a standard batch size of 4 and a learning rate of 5e-4

7 Future Works

Our current approach lacks data augmentation during training, primarily due to the utilization of pretrained models trained on H&E images, which inherently minimizes the domain gap. Our forthcoming step involves retraining these models while introducing image augmentation techniques to potentially enhance their generalizability and performance.

Additionally, we didn't conduct an extensive hyperparameter tuning for our models. Exploring and optimizing the hyperparameters holds promise for improving the performance of all models assessed in our study. This refinement process could potentially unveil further enhancements in their efficacy and predictive capabilities.

References

- G. Bradski. The opencv library. *Dr. Dobb's Journal of Software Tools*, 2000.
- Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.*, 9:62–66, 1979. URL <https://api.semanticscholar.org/CorpusID:15326934>.
- Yi Zheng, Rushin Gindra, Emily Green, Eric Burks, Margrit Betke, Jennifer Beane, and Vijaya Kolachalama. A graph-transformer for whole slide im-

age classification. *IEEE transactions on medical imaging*, PP, 05 2022. doi: 10.1109/TMI.2022.3176598.

Ming Lu, Drew Williamson, Tiffany Chen, Richard Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature Biomedical Engineering*, 5:1–16, 06 2021. doi: 10.1038/s41551-020-00682-w.

Xiyue Wang, Sen Yang, Jun Zhang, Minghui Wang, Jing Zhang, Wei Yang, Junzhou Huang, and Xiao Han. Transformer-based unsupervised contrastive learning for histopathological image classification. *Medical Image Analysis*, 81:102559, 2022. ISSN 1361-8415. doi: <https://doi.org/10.1016/j.media.2022.102559>. URL <https://www.sciencedirect.com/science/article/pii/S1361841522002043>.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

Maximilian Ilse, Jakub M. Tomczak, and Max Welling. Attention-based deep multiple instance learning, 2018.