

# Digital Pathology-based Study of Cell- and Tissue-level Morphologic Features in Serous Borderline Ovarian Tumor and High-grade Serous Ovarian Cancer

Jun Jiang<sup>1\*</sup>, Burak Tekin<sup>2\*</sup>, Ruifeng Guo<sup>2</sup>, Hongfang Liu<sup>1</sup>, Yajue Huang<sup>2</sup>, Chen Wang<sup>3</sup>

<sup>1</sup>Department of Artificial Intelligence and Informatics, Mayo Clinic, Rochester, MN, USA, <sup>2</sup>Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA, <sup>3</sup>Department of Health Science Research, Mayo Clinic, Rochester, MN, USA

\*Dr. Jiang and Dr. Tekin contribute the same to this work

Submitted: 12-Sep-2020

Revised: 28-Dec-2020

Accepted: 11-Feb-2021

Published: 05-Jun-2021

## Abstract

**Background:** Serous borderline ovarian tumor (SBOT) and high-grade serous ovarian cancer (HGSOC) are two distinct subtypes of epithelial ovarian tumors, with markedly different biologic background, behavior, prognosis, and treatment. However, the histologic diagnosis of serous ovarian tumors can be subjectively variable and labor-intensive as multiple tumor slides/blocks need to be thoroughly examined to search for these features. **Materials and Methods:** We developed a novel informatics system to facilitate objective and scalable diagnosis screening for SBOT and HGSOC. The system was built upon Groovy scripts and QuPath to enable interactive annotation and data exchange. **Results:** The system was used to successfully detect cellular boundaries and extract an expanded set of cellular features representing cell- and tissue-level characteristics. The performance of cell-level classification for both tumor and stroma cells achieved >90% accuracy. The performance of differentiating HGSOC versus SBOT achieved 91%–95% accuracy for 6485 imaging patches which have sufficient tumor and stroma cells (minimum of ten each) and 97% accuracy for classifying patients when aggregating the results to whole-slide image based on consensus. **Conclusions:** Cellular features digitally extracted from pathological images can be used for cell classification and SBOT v. HGSOC differentiation. Introducing digital pathology into ovarian cancer research could be beneficial to discover potential clinical implications. A larger cohort is required to further evaluate the system.

**Keywords:** Digital pathology, high-grade serous ovarian cancer, serous borderline ovarian tumor

## INTRODUCTION

Ovarian cancer is one of the leading causes of death in patients with gynecological malignancies. Epithelial ovarian tumors are classified according to the histology subtypes, with the serous type being the most common, and further divided into benign, borderline, and malignant (carcinoma) categories. Serous borderline ovarian tumors (SBOT) represent approximately 5%–10% of all ovarian serous tumors.<sup>[1]</sup> Compared to benign ovarian tumors, borderline ovarian tumors exhibit greater epithelial proliferation and cellular atypia. However, in contrast to their malignant counterparts, borderline ovarian tumors lack destructive stromal invasion.<sup>[2]</sup> SBOTs share genetic changes with low-grade serous ovarian carcinomas (LGSOs), for example, *KRAS* and *BRAF* mutations, and can progress to the latter in a subset of patients. Overall, SBOTs are associated

with a favorable prognosis, with 5-year survival rates of early-stage patients as high as 90%.<sup>[1]</sup> High-grade serous ovarian cancer (HGSOC), on the other hand, is a biologically distinct entity more commonly related to *TP53* mutations, most often presenting at an advanced stage and associated with a relatively poor prognosis.<sup>[3]</sup> Thus, the histologic distinction

**Address for correspondence:** Dr. Chen Wang,  
Mayo Clinic 200 First St. SW Rochester, MN 55905, USA.  
E-mail: wang.chen@mayo.edu  
Dr. Yajue Huang,  
Mayo Clinic 200 First St. SW Rochester, MN 55905, USA.  
E-mail: huang.yajue@mayo.edu

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

**For reprints contact:** WKHLRPMedknow\_reprints@wolterskluwer.com

**How to cite this article:** Jiang J, Tekin B, Guo R, Liu H, Huang Y, Wang C. Digital pathology-based study of cell- and tissue-level morphologic features in serous borderline ovarian tumor and high-grade serous ovarian cancer. *J Pathol Inform* 2021;12:24.

Available FREE in open access from: <http://www.jpathinformatics.org/text.asp?2021/12/1/24/317823>

### Access this article online

#### Quick Response Code:



**Website:**  
[www.jpathinformatics.org](http://www.jpathinformatics.org)

**DOI:**  
10.4103/jpi.jpi\_76\_20

between SBOTs and HGSOCS is important in that it has major prognostic and therapeutic implications.<sup>[4]</sup> The histologic diagnosis of HGSOCS heavily relies on the morphologic assessment of tumor cells, and the presence of marked variation in nuclear size and shape represents an important feature pointing to the diagnosis of HGSOCS. In general, SBOTs can be differentiated from LGSOCS and HGSOCS by lack of destructive stromal invasion and high-grade morphologic features, respectively. However, SBOTs can display “micro-invasion” and high-grade morphologic features may be very focal in a given HGSOCS case, potentially posing diagnostic challenges and necessitating extensive evaluation of multiple surgical blocks of resected specimens.<sup>[3]</sup> Overall, the accurate histologic diagnosis of serous ovarian tumors can be labor-intensive and subjective.

With the advancement of digital pathology (dPath), there have been substantial research interests in developing image analysis approaches for characterizing tissue heterogeneity digitally in various cancer types and studying associations with clinical outcomes. For examples, Lu *et al.*<sup>[5]</sup> found that the nuclear shape and orientation features are predictive of survival of early-stage estrogen receptor-positive breast cancers. Lan *et al.*<sup>[6]</sup> showed that quantitative measurements of the extent and density of lymphocytic infiltration was significantly associated with overall survival and progression-free survival in ovarian cancers. Beck *et al.*<sup>[7]</sup> showed that nuclear morphologic features within the stroma were significantly associated with survival in breast cancer. Nawaz *et al.*<sup>[8]</sup> revealed that tumor spatial heterogeneity was a strong prognostic factor in ovarian cancers. Moreover, dPath approaches were also developed for facilitating diagnoses of different tumor subtypes within a single cancer type. For examples, Zhang *et al.*<sup>[9]</sup> proposed a stepwise method to classify Non-Hodgkin’s lymphoma subtypes, over 99% cross validation accuracy was achieved on differentiating chronic lymphocytic leukemia, follicular lymphoma, and mantle cell lymphoma. Rathore *et al.*<sup>[10]</sup> demonstrated that deep learning techniques are capable of predicting overall survival and molecular markers (isocitrate dehydrogenase gene mutation and co-deletion of chromosomes 1p and 19q) in gliomas. Barker *et al.*<sup>[11]</sup> proposed a brain tumor subtype classification model; spatially localized features of shape, color, and texture from whole-slide image (WSI) tiles were eligible in differentiating glioblastoma multiforme and lower grade glioma.

Although SBOT and HGSOCS taxonomically belong to the same umbrella category of ovarian epithelial tumors, they are markedly distinct entities with dissimilar biologic behavior and histologic findings. Based on the latter, it can be anticipated that they have different cellular composition and spatial heterogeneity. Motivated by the aforementioned diagnostic challenges (subjective and labor-intensive histopathologic examinations), we sought to investigate dPath feasibility of classifying HGSOCS and SBOT from cellular and tissue levels using machine learning (ML) approaches. In order to extract labels and features for ML purposes, a series of informatics

modules were designed, including interactive pathological annotations, objective quantifications of imaging features, and scalable computations to thoroughly scan entire WSI. The overall flow of this developed informatics process is shown in Figure 1. First, interactive pathology annotation and machine abstraction were completed through QuPath-Groovy data communication. Then, cellular features were extracted to train ML predictive models to classify cell types, and tissue-level histology classifications were done by aggregating cellular level features. Overall, accurate prediction results from cellular and tissue levels were successfully achieved, strongly suggesting great potentials for applying dPath methods to broad research and clinical applications.

## MATERIALS AND METHODS

### Cohort selection

Thirty cases (15 cases of SBOT and 15 cases of HGSOCS) with unequivocal histologic features were randomly retrieved from the institutional pathology system database. The diagnoses were independently confirmed by two pathologists (R.G. and Y.H.). Clinical features of the cases are outlined in Table 1. The Regional Institutional Review Board approved the study.

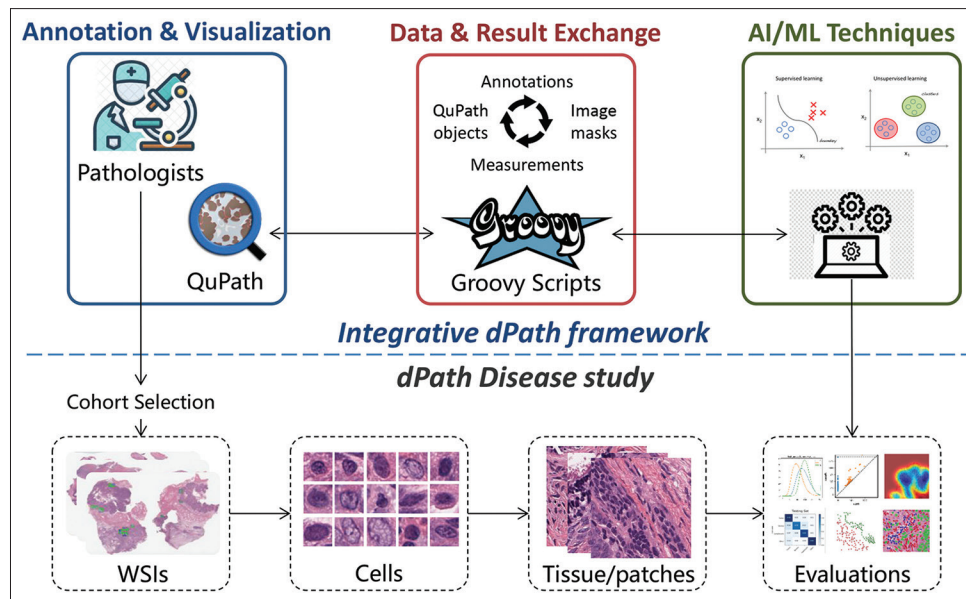
### Image acquisition and annotation

Archived slides were scanned at  $\times 40$  using the digital whole-slide scanner (Aperio Scanscope XT), with  $0.25 \mu\text{m} \times 0.25 \mu\text{m}$  of pixel size in obtained WSIs.

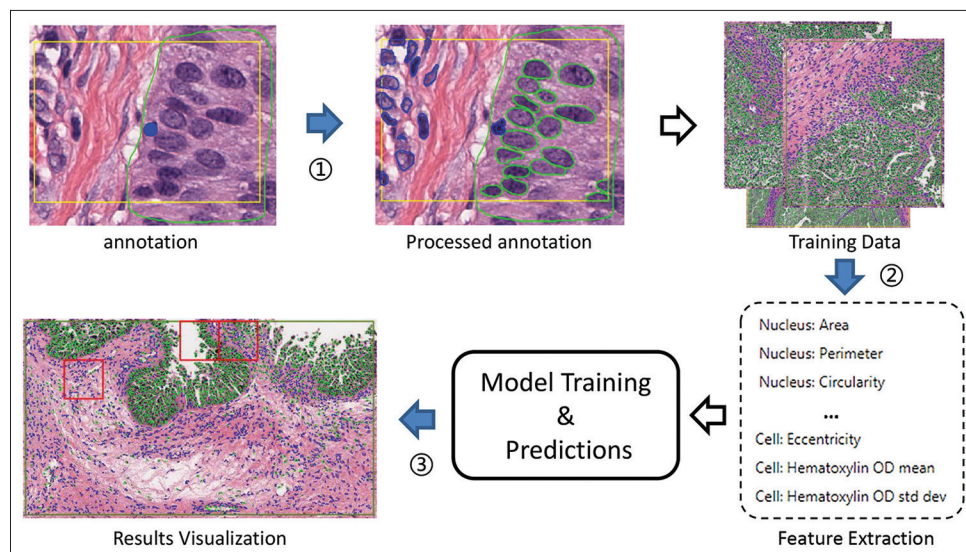
A training set, consisting of ten WSIs (5 each for SBOT and HGSOCS), was annotated by two pathologists. In each case, five regions most representative of SBOT/HGSOCS morphology features were chosen as regions of interest (ROI) for manual annotations, with each ROI being larger than  $256 \mu\text{m} \times 256 \mu\text{m}$ . Cells in the ROIs were annotated as either tumor or stroma cells. Rather than directly delineating the boundary of each cell and then assigning labels to each cell, we proposed a simplified annotation process in which pathologists used polygons to outline homogeneous regions of similar cells, except those deliberately annotated by points. Based on built-in interfaces in QuPath, we made customized scripts to process pathologists’ annotations (details seen in section below).

### Groovy-based interactive pathology annotation and data abstraction

As a powerful tool for quantitative pathology and bio-image analyses, QuPath provides application programming interface to enable high-throughput analysis across many images. This greatly extends the feasibility of customizing pathological image processing pipelines. Groovy was recommended by QuPath as the best programming language to use for image processing, interactive annotation and visualization, since it closely matches the Java programming language in which the majority of QuPath itself was written. In order to build a dPath framework with flexible pathology annotations and expandable ML modules, a Groovy-based interaction middleware was designed and implemented to communicate



**Figure 1:** Overall informatics framework for digital pathology analysis of serous borderline ovarian tumor versus HGSOc differentiation



**Figure 2:** Relevant major steps for Groovy-based annotation and data abstraction. Each blue arrow denotes a Groovy middleware, including ① annotation processing, ② feature extraction and ③ result visualization

between QuPath and Python (or any programming languages performing ML tasks).

1. Annotation module [Figure 2①]. The main objective of this module was to reduce the manual annotation of outlining cellular boundaries. Homogeneous regions and points annotated by pathologists were processed with the aid of customized scripts and a built-in interface in QuPath. Within this process, cells were detected using a watershed segmentation plugin, and category labels from annotations (homogeneous regions and points) were then assigned to individual cells. Parameters of this cell segmentation procedure were fine-tuned according to training dataset: pixel size was set to 0.25  $\mu\text{m}$ , minimum nuclear area was set to 10  $\mu\text{m}^2$

2. Cellular feature extraction module [Figure 2]. After cell segmentations, cellular features were extracted from SBOT and HGSOc cases according to morphological and color-intensity characteristics. These features were constructed according to different cellular components, i.e., nuclear, cytoplasm and cell. Nuclei were automatically segmented from the background using a watershed nuclear segmentation method. The boundaries of nuclei were arbitrarily expanded up to 5  $\mu\text{m}$  or until the expansions overlapped with adjacent cells. The extended areas were regarded as cytoplasm. Cells were identified as the integration of nuclear and cytoplasm components. Morphological features were calculated based on the binary mask of each nucleus/cytoplasm/cell, which



**Table 1: Clinical characteristics of the study cases**

Case number	Age at diagnosis (years)	Year of diagnosis	Stage at diagnosis
HGSOC			
1	69	2011	IC1
2	51	2015	IIIC
3	64	2011	IIIC
4	48	2004	IIIC
5	86	2004	IIIC
6	73	2002	IIIC
7	77	2009	IIIC
8	47	2008	IIIC
9	76	2007	IIIC
10	69	2010	IIIC
11	48	2012	IIIC
12	60	2006	IC3
13	86	2001	IIIC
14	55	2012	IC1
15	71	2003	IC3
SBOT			
16	80	2006	IA
17	20	2001	IA
18	67	2009	IA
19	71	2002	IA
20	35	2002	IIIC
21	38	2003	IA
22	46	2014	IA
23	69	2005	IIIC
24	28	2004	IIIC
25	29	2006	IC3
26	60	2002	IIB
27	64	1999	IIC
28	59	2012	IA
29	64	2001	IIIA2
30	46	2001	IA

\*Limited clinical follow-up information available. Staging based on (The International Federation of Gynecology and Obstetrics) FIGO's staging classification for cancer of the ovary, fallopian tube, and peritoneum (2014) (Jaime Prat - corresponding author and FIGO Committee on Gynecologic Oncology; J Gynecol Oncol. 2015 Apr; 26 (2): 87-89.)

were used to describe the geometry properties of the cells including area, perimeter, etc., In order to enrich color-intensity characteristics of nucleus/cytoplasm/cell, original Hematoxylin and Eosin (H&E) images were decoupled into H and E components. Color-intensity features were calculated based on the image attributes under the binary mask of each cell, including H and E optical density (OD) mean, standard deviation, etc., With customized scripts, all the features were calculated in QuPath and exported into csv files, in which each row is a sample of a cell, the first column is the annotated label of the cell, and the remaining columns are feature values

- Visualization module [Figure 2]. Cell/tissue classification and some intermediate results (such as binary images, polygon coordinates etc.,) were converted into QuPath

objects and imported into QuPath for convenient result visualization and examination.

In summary, Groovy exports pathologist's manual annotations, such as cellular coordinates in images and cell-type labels into a dataset for building predictive ML models. Once a predictive model is built, cellular and tissue-level results can be returned by the ML module and displayed in QuPath through coordination of the Groovy communication module. This enables further examination of misclassified cells and tissue regions, for both quality assurances and iterative improvement of model training processes. All the source codes and relevant documentations are made available to the public in GitHub repository (<https://github.com/smujiang/CellularComposition>).

## Image analysis

### Cellular classification and examinations of feature importance

With  $n = 41$  extracted features, linear support vector machine (SVM) was introduced to classify cells and clarify the importance of features in cell type identification. Each cell in SVM was interpreted as a sample in the virtual space spanned by all the features, and the algorithm was trained to fit a hyperplane (decision boundary) aiming to maximize the margin between sample types. Support vectors were chosen as most representative samples (cells) to determine the decision boundary. Distances from sample to hyperplane were abstracted to reflect the likelihood of a sample to be correctly classified. Once the hyperplane was determined, the coefficients of the trained model were used to determine feature importance of cell classification, since the weights account for the orthogonal vector coordinates orthogonal to the hyperplane and their direction represents the predicted class. A line chart of feature weights was used to show the difference in feature importance for cell classification tasks in SBOT and HGSOC. A confusion matrix was calculated to examine cell classification accuracy.

### Tissue patch classification for serous borderline ovarian tumor versus HGSOC differentiation

For each case (15 SBOTs and 15 HGSOCs in total), 10 ROIs with both tumor and stroma areas were selected for classification evaluation purposes. The cell classification model trained in the previous phase was applied to these ROIs to differentiate tumor and stroma cells. ROIs were divided into regular  $512 \times 512$  pixel patches to enable measurements of local differences. Patches that contained at least 10 tumor and 10 stroma cells were deemed as eligible patches for ML purposes. For each eligible patch, total  $n = 609$  features were aggregated from cell-level results, including (1) statistics of cellular features: summarizing from cellular features, seven patch-level features were constructed according to distributions: mean, median, standard deviations, (Q1, Q3), and minimum and maximum values per cellular feature type and cell-type. (2) Tumor-stroma interaction features. As tumor cells tend to group into clusters, the Gaussian-based kernel

density estimation (KDE) was used to fit an empirical density distribution of tumor cells within a patch. Then, this fitted KDE function was used to evaluate relative distance of stromal cells with respect to tumor cell clusters. In order to capture tumor-stroma interaction from multiple scales, KDE kernel width was set to 16, 20, 24, 30, and 34. For each stromal cell, we calculated a probability/likelihood score based on KDE, and seven statistics related to distributions of the scores were included in our patch descriptor.

Considering the high dimensionality of patch-level features, a regression method with sparsity penalty, named least absolute shrinkage and selection operator (LASSO), was used to highlight the most important features contributing to patch-level predictions. In LASSO cost function, the penalty term normalizes the coefficients such that the coefficients that take large values get penalized, resulting in shrinking the count of non-zero coefficients which helps to reduce the model complexity and multi-collinearity.

## RESULTS

### Cell classification

Within training set, 17,181 tumor cells and 8828 stroma cells were annotated for HGSOC cases (on average 3436 tumor cells and 1766 stromal cells per WSI), and 2638 tumor cells and 6435 stroma cells for SBOT cases (on average 527 tumor cells and 1287 stromal cells per WSI). Using SVM classifier based on 41 cellular features, 86.4%–89.1% cell classification accuracies were achieved in HGSOC, and 85.4%–90.8% in SBOT cases [Figure 3a and b]. Eosin OD intensity was found to play the leading role in differentiating stroma and tumor cells in both categories [Figure 3c]. Unsupervised clustering of feature dependences demonstrated that features with similar morphology and intensity implications were often highly correlated [Supplementary Figure 1].

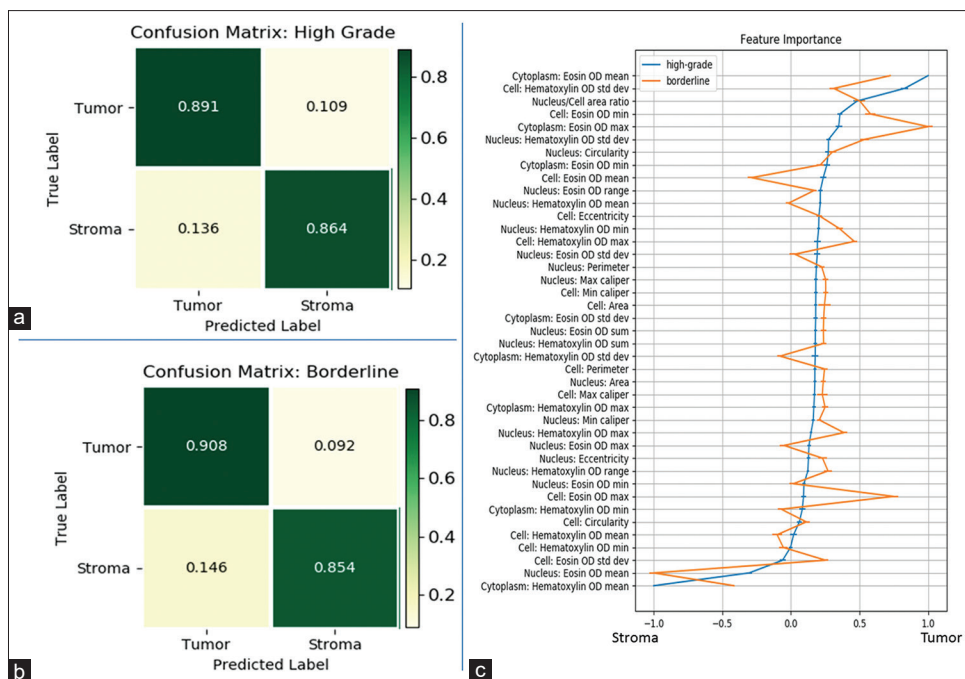
In order to identify cell-level misclassifications, K-nearest neighbors algorithm ( $K = 10$ ) was used to cluster 3575 misclassified cells. Through manual examinations, 44.2% (1580) could be attributed to explainable process errors, such as under/over cellular segmentation, histologic artifacts induced by slide preparation/scanning, or nonspecific cellular/tissue elements that could not be unequivocally identified as either tumor or stroma cells (e.g., portions of adipocytes or air/fat bubbles, heavily pigment laden cells, red blood cells/hemorrhage, or possible non-cellular connective tissue fragments) [Supplementary Figure 2]. Besides examining individually misclassified cells, investigation was conducted to examine whether some support vectors, as representative cells for SVM classifier, may contribute to systematic classification errors. Similar to the misclassified cells, the classification of cells associated with support vectors was complicated by the presence of the aforementioned nonspecific elements, which indicates that elaborative features for these cells are essential for performance improvement. Of note, when taken out of tissue context, a subset of misclassified cells was challenging for the

pathologists to confidently classify as either tumor or stroma cells, which indicates potential morphologic variability between different cases and reflects the pathologists' practice of evaluating the cells in light of the totality of the slide in a given case.

### Tissue- and subject-level differentiations of serous borderline ovarian tumor versus HGSOC

In order to conduct tissue- and subject-level classifications, 300 ROIs (150 HGSOC vs., 150 SBOT) were selected with cell detection and classification, as previously described (image analysis section). In total, 903,678 cells were detected from HGSOC ROIs, 404,973 of which were tumor cells; 465,299 cells were detected from SBOT ROIs, 151,393 of which were tumor cells. For evaluating tissue-level discrimination between SBOT and HGSOC, ROIs were divided into multiple regular image patches of  $512 \times 512$  [Figure 4a and b]. In total, 6446 and 4025 image patches were extracted from HGSOC and SBOT cases, respectively. Based on cellular features and tumor-stroma spatial distributions, 609 patch-level features were constructed for each image patch. In particular, kernel width of KDE was varied to provide multi-resolution features for characterizing tumor-stroma reaction. In total 6485 (HGSOC: 4225, SBOT: 2260) eligible image patches were selected from 10,471 (HGSOC: 6446, SBOT: 4025) patches as eligible for tissue (patch) classification. With 609 dimensional features, SVM was trained to differentiate SBOT and HGSOC and overall patch-level accuracies were 90.5%–90.7% [Figure 4c]. In order to evaluate classification separations, distances of image patches to SVM classification hyper-plane were computed. Further, distances per histotype histograms demonstrated that aggregation of cellular features has largely separated HGSOC versus SBOT [Figure 4d]. To further delineate features with the most critical contributions to classification performance, sparsity-based LASSO regression was applied revealing 15 features with nonzero coefficients [Table 2]. We found that geometry features, such as cell area and perimeter, did not play a significant role in cell classification, but were important for tissue differentiation. Statistical values from hematoxylin intensity of tumor and stroma were strongly associated with SBOT v. HGSOC classification. KDE features reflective of stroma-tumor interactions did not contribute significantly for SBOT v. HGSOC differentiation. Moreover, when aggregating scores from multiple tissue patches into subject-level with bootstrap resampling, 97% (29/30) accuracy was achieved [Figure 4e], which shows that tissue-level features were potentially significant for subject classification.

Through histopathologic reviews of misclassified patches ( $n = 199$ ), several representative morphologic features were found in multiple patches and cases. For example, (i) some patches from HGSOC had compact populations of cells with nearly overlapping cellular borders, which might have arguably influenced accurate segmentation of cells. (ii) Some HGSOC cases had a predominance of cells with an optically clear nuclear and/or cytoplasmic



**Figure 3:** Cellular-level classification results. (a and b) Confusion matrix of cell classification in HGSOC and serous borderline ovarian tumor. (c) Feature ranking based on support vector machine classifier. Feature importance for cell classification in high grade and borderline cases. Feature importance (X axis) was normalized to (−1, 1). Ticks of Y axis are the features (sorted by the importance of HGSOC cases) selected by support vector machine for tumor and stroma classification

appearance. A possible explanation for the misclassification of these patches may be related to their overall lower eosin or hematoxylin intensity compared to other HGSOC cases with more hyperchromatic (darker) nuclei. (iii) Another common feature among misclassified patches was the presence of a spindle-cell population aligned in a streaming fashion [Supplementary Figure 3].

## DISCUSSION

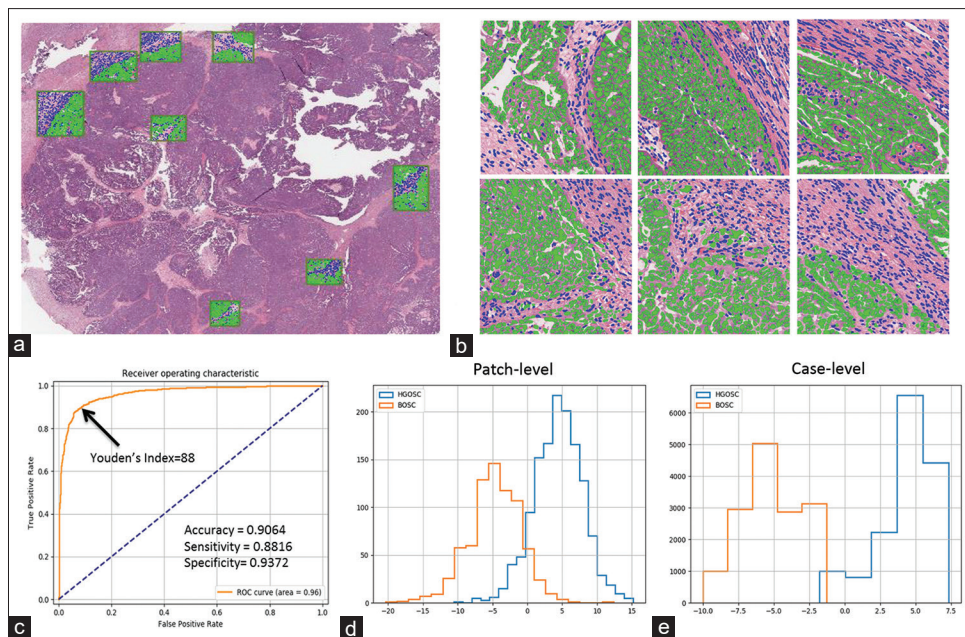
Although abundant work has been done on cell segmentation and classification,<sup>[12-14]</sup> one of the key challenges is over-and under-segmentation due to high variation of shapes and textures.<sup>[12]</sup> These challenges were also found in this work. An automatic cell segmentation method (Watershed Segmentation) was introduced into our work to eliminate the need for pathologists to directly delineate cell borders in the annotation step. According to this study, over-and under-segmentation were more likely to be observed in tumor cells. This may be attributable to greater variation of both shape and texture of tumor cell nuclei in comparison to stromal cell nuclei. As a contrast, lymphocytes were less likely to be over-or under-segmented due to their relatively uniform morphology. The consequence of inaccurate segmentation results will propagate errors to the following steps, including morphological and texture feature extraction, cell classification model training, etc., For example, one cell may be over-segmented into several parts, while two adjacent cells without clear boundaries may be under-segmented into

one cell. Both instances lead to statistical errors in calculating the cell radius.

Several features in this study deserve further mention from the pathologists’ perspective. The cellular features evaluated herein overlap with some of the parameters that are routinely assessed in the surgical pathology practice, such as nuclear-to-cytoplasmic ratio or the presence of hyperchromatic nuclei. However, the algorithm of this study in its current form does not address two features that are frequently evaluated by pathologists when differentiating benign and malignant lesions, namely pleomorphism and detection of nucleoli. The former refers to size and shape variation of cells and the latter refers to small spherical structures in the nucleus. The presence of marked pleomorphism or prominent nucleoli may favor malignancy, within the appropriate context. Supplementary Figure 3 shows examples of misclassified patches from HGSOC cases that, upon review, were observed to have pleomorphism and/or prominent nucleoli. The inclusion of these features into the analysis can further refine the classification accuracy in follow-up studies.

In pathologists’ daily practice, the final diagnosis represents an overall and somewhat subjective interpretation of many considerations, including clinical factors such as patient’s age, size or growth rate of the tumor, as well as morphologic features on the slide, and ancillary studies such as immunohistochemical stains. As such, it remains challenging for constructing mathematical model to approximate the human diagnostic thinking process. Of note, a given histologic slide





**Figure 4:** Tissue- and subject-level classification results. (a) Cell detection and classification were conducted in several regions of interests per whole slide image. (b) Regions of interests with cell labels were divided into regular image patches. (c) Receiver operating characteristic curve of patch classification with aggregated cellular features. (d) Tissue-level histograms of distances to support vector machine hyper-plane. (e) Subject-level aggregated distances after  $n = 1000$  bootstrapping

**Table 2: Important features selected by least absolute shrinkage and selection operator classifier**

Content	Cellular feature name	Statistics	Importance
stroma_Cell	Eosin OD minimum	Minimum	0.11
stroma_Cell	Hematoxylin OD maximum	Q3	0.33
stroma_Nucleus	Hematoxylin OD maximum	Minimum	-0.33
stroma_Nucleus	Hematoxylin OD maximum	Q1	-0.35
tumor_Cell	Area	Mean	0.1
tumor_Cell	Eosin OD maximum	Maximum	0.21
tumor_Cell	Eosin OD maximum	Q3	0.23
tumor_Cell	Hematoxylin OD maximum	Q3	0.14
tumor_Cytoplasm	Hematoxylin OD maximum	Median	-0.28
tumor_Cytoplasm	Hematoxylin OD mean	Maximum	-0.9
tumor_Cytoplasm	Hematoxylin OD maximum	Q3	-0.24
tumor_Cytoplasm	Eosin OD maximum	Maximum	0.26
tumor_Nucleus	Perimeter	Mean	-0.12
tumor_Nucleus	Hematoxylin OD maximum	Maximum	0.17
tumor_Nucleus	Hematoxylin OD mean	Q3	-0.16

OD: Optical density

may include areas of varying morphologic characteristics, which was also observed in this study. What further complicates development of mathematical models and dPath algorithms is that, on a case-by-case basis, certain morphologic features can potentially “trump” or “overrule” others from the pathologists’ perspective. The only misclassified case in the study (Case 12) represents an example for this phenomenon in that it harbors large cells with bizarre, hyperchromatic nuclei and a high nuclear-cytoplasmic ratio. A review of the misclassified patches of this case (75 misclassified out 132

eligible patches) showed that these cells were outnumbered by the surrounding, more uniform population of smaller cells, which might have skewed the algorithm towards misclassifying these patches as borderline. However, in clinical pathology practice, the presence of the large cells with bizarre, hyperchromatic nuclei even if few in number would “trump” other morphologic features and make the pathologist favor a diagnosis of HGSOC over SBOT. Another noteworthy feature in this misclassified case is the presence of prominent nucleoli in a subset of cells [Figure 5]. Further refinement of the study algorithm to incorporate the evaluation of “outlier” cells with remarkably different, bizarre sizes and shapes, overall cellular pleomorphism, and prominent nucleoli would improve classification accuracy, especially in cases like Case 12 [Table 1].

In recent years, deep learning based approaches were employed for cell segmentation and classification in many cancer research tasks.<sup>[13-15]</sup> These advanced methods are considered to be superior to traditional ML methods, as cell classification model for one cancer could be easily retrained with transfer learning strategy for another cancer. However, morphologic features across cancers and their subtypes could be dramatically different. Especially for ovarian cancer, since it’s is not a homogeneous disease, but rather a group of diseases-each with different morphology and biological behavior.<sup>[16]</sup> This means that large numbers of cells need to be annotated for specific cancers due to the data-hungry nature of these deep learning models, making cell annotation impractical for many application scenarios. Even if deep neural networks are trained for a specific task, interpretability

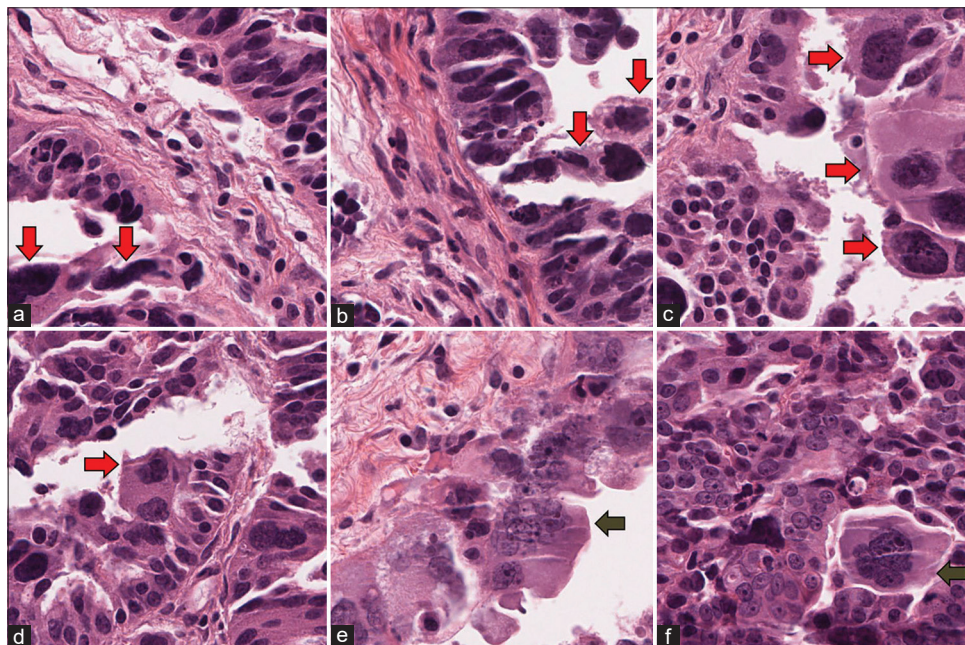
is often limited. In deep learning models, it is hard to explain which elements play a more important role in downstream analysis because feature extraction components are embedded in multiple layers of the networks. Researchers have to use step-wise methods (detect individual cells, extract features for each cell, and classify cells according to the features) to evaluate the significance of a specific feature and connect this information to molecular/genomic discoveries. In this work, we constructed cellular features according to both morphological and color-intensity characteristics, and trained the model in a relatively small dataset. The encouraging evaluation results suggested that our approach reconciled classification capability and feature interpretability. Leveraging interactive annotation pipeline and collaborative pathologists, we were able to obtain a considerable number of cells and patches for model training and evaluation. However, more extensive case level assessments are still needed to consolidate predictive models truly applicable to clinics.

As future methodology improvements, at least two aspects were worthy mentions: (i) stromal invasion: In clinical practice, two main histologic features are utilized to differentiate between SBOT and HGSO, namely stromal invasion and cellular morphology. This study with its current methodology, however, does not address the stromal invasion, but mainly focuses on the evaluation of morphologic features. This represents a limitation, and we are planning future studies investigating stromal invasion. Nevertheless, this point could also be perceived as a relative strength since solely using

the cellular morphologic features, our current algorithm was able to accurately classify the vast majority of the cases. (ii) inclusion of more histological features: In this work, decoupled H and E intensity information played an important role in cell classification, even if the cells were not well segmented. In practice, a limitation of employing H and E intensity for classification purposes is that this parameter depends on multiple factors, such as the histologic staining process or the duration since the initial preparation of slides as the latter typically tend to fade with time. Despite these limitations, our cell classification accuracy reached up to 87.8%. Our further work will put more efforts on developing cell segmentation and classification methods with better performance, and incorporating domain-specific histologic features such as pleomorphism and prominent nucleoli into the analysis.

## CONCLUSIONS

In this study, we developed an interactive informatics system to digitally annotate areas, construct features and classify cell types, focusing on a case study of differentiating borderline and high-grade serous ovarian tumors. Within the developed digital pathology framework, a number of cellular and sub-cellular features were aggregated to tissue regional and whole-slide level for achieving accurate histology classification. Through close examinations of machine-classified cellular and regional results, we revealed pathologically interpretable features, as well as unexpected findings after pathology re-review, such as an atypical case with unusual morphologic



**Figure 5:** Image patches from case 12, the only misclassified case in the study. There are multiple cells with nondescript, *bizarre* shapes and hyperchromatic nuclei (red arrows). In practice, the presence of these cells, even in the absence of other worrisome features, would significantly raise the level of the pathologists' concern about a high-grade malignancy. Some of the cells in (c and d) appear to have two nuclei. (e and f) Captured from a different area of the whole slide image, demonstrate a strikingly different morphology, with a distinct population of cells displaying less hyperchromatic nuclei with discernible nucleoli. Some of the cells appear merged together, forming giant cells with a somewhat *syncytial* appearance (green arrows) (H&E). The image patches in each panel are 512 pixels  $\times$  512 pixels and approximately correspond to  $\times$  200 magnification



features. Together with a developed scalable system enabling interactive annotations and advanced machine learning methods, data-driven findings as such will warrant large cohort to systematically investigate associations with pathological results and clinical outcomes.

### Acknowledgment

This research was partially supported by the Mayo Clinic Ovarian Cancer SPORE (P50 CA136393).

### Financial support and sponsorship

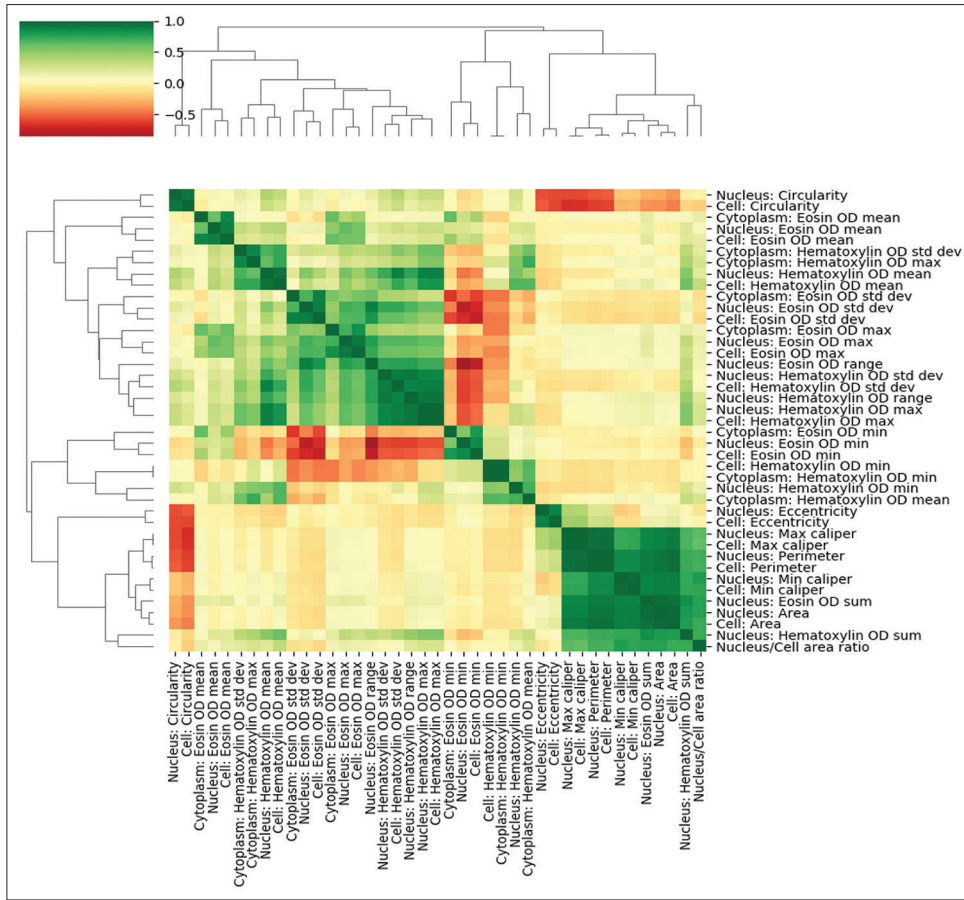
Mayo Clinic Ovarian Cancer SPORE (P50 CA136393).

### Conflicts of interest

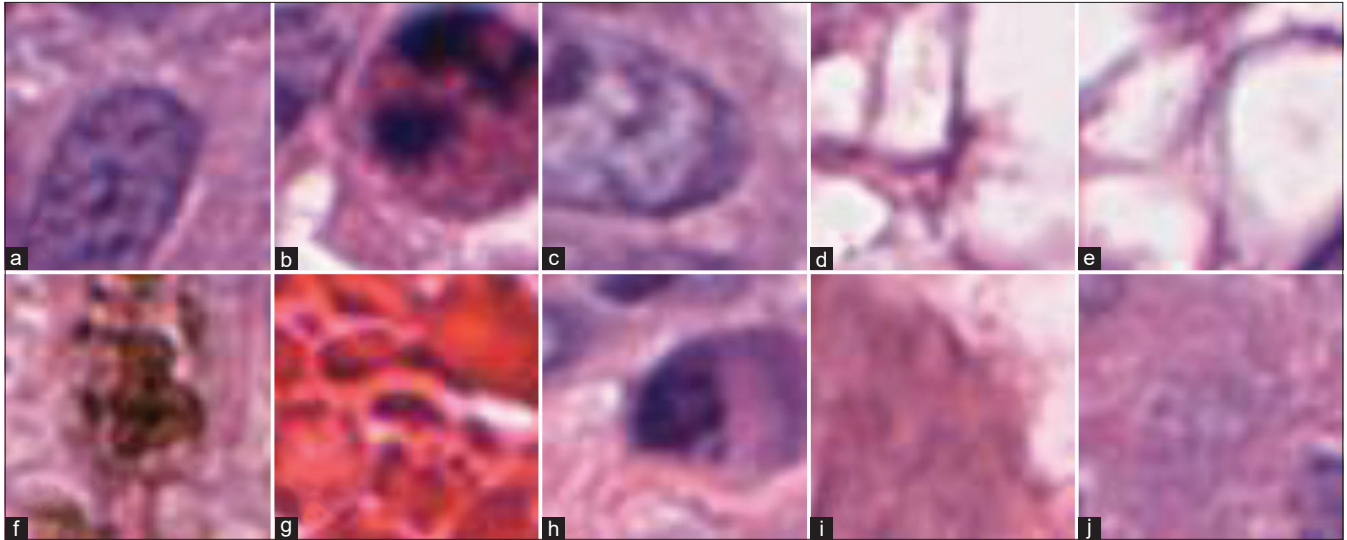
There are no conflicts of interest.

## REFERENCES

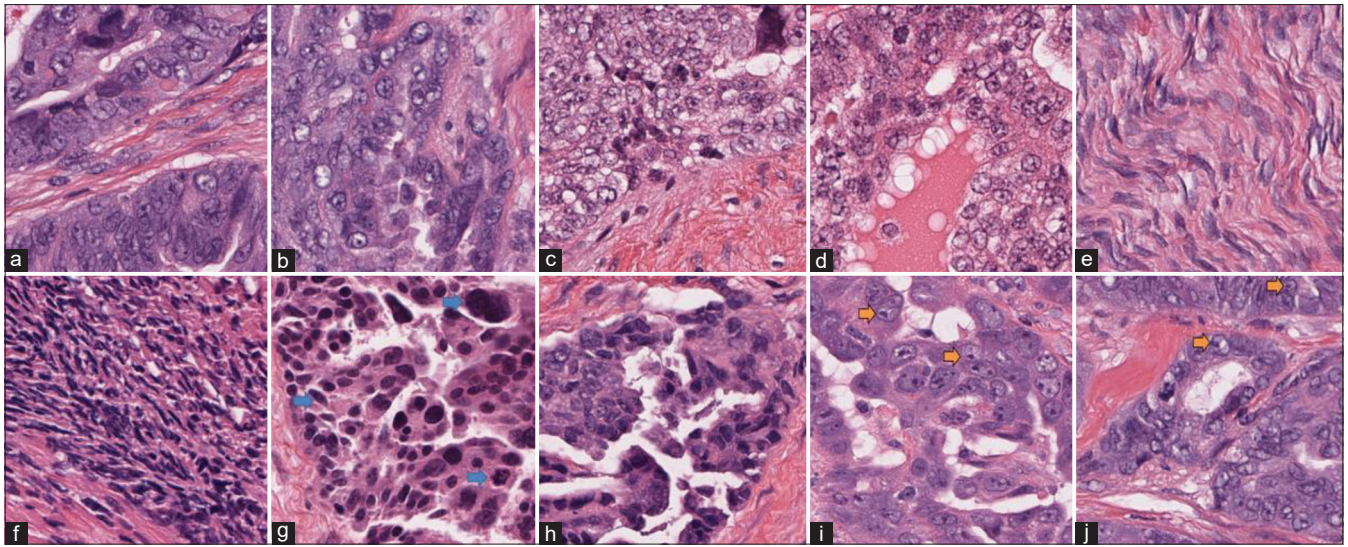
- Prat J. Pathology of borderline and invasive cancers. *Best Pract Res Clin Obstet Gynaecol* 2017;41:15-30.
- Hannibal CG, Frederiksen K, Vang R, Kurman RJ, Kjaer SK. Risk of specific types of ovarian cancer after borderline ovarian tumors in Denmark: A nationwide study. *Int J Cancer* 2020;147:990-5.
- Hauptmann S, Friedrich K, Redline R, Avril S. Ovarian borderline tumors in the 2014 WHO classification: Evolving concepts and diagnostic criteria. *Virchows Arch* 2017;470:125-42.
- Hacker KE, Uppal S, Johnston C. Principles of treatment for borderline, micropapillary serous, and low-grade ovarian cancer. *J Natl Compr Canc Netw* 2016;14:1175-82.
- Lu C, Romo-Bucheli D, Wang X, Janowczyk A, Ganesan S, Gilmore H, *et al.* Nuclear shape and orientation features from H and E images predict survival in early-stage estrogen receptor-positive breast cancers. *Lab Invest* 2018;98:1438-48.
- Lan C, Heindl A, Huang X, Xi S, Banerjee S, Liu J, *et al.* Quantitative histology analysis of the ovarian tumour microenvironment. *Sci Rep* 2015;5:16317.
- Beck AH, Sangoi AR, Leung S, Marinelli RJ, Nielsen TO, van de Vijver MJ, *et al.* Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci Transl Med* 2011;3:108ra113.
- Nawaz S, Trahearn NA, Heindl A, Banerjee S, Maley CC, Sottoriva A, *et al.* Analysis of tumour ecological balance reveals resource-dependent adaptive strategies of ovarian cancer. *EBioMedicine* 2019;48:224-35.
- Zhang J, Cui W, Guo X, Wang B, Wang Z. Classification of digital pathological images of non-Hodgkin's lymphoma subtypes based on the fusion of transfer learning and principal component analysis. *Med Phys* 2020;47:4241-53.
- Rathore S, Iftikhar MA, Mourelatos Z. Prediction of overall survival and molecular markers in Gliomas via analysis of digital pathology images using deep learning. *ArXiv* 2019;9:124.
- Barker J, Hoogi A, Depeursinge A, Rubin DL. Automated classification of brain tumor type in whole-slide digital pathology images using local representative tiles. *Med Image Anal* 2016;30:60-71.
- Aeffner F, Zarella MD, Buchbinder N, Bui MM, Goodman MR, Hartman DJ, *et al.* Introduction to digital image analysis in whole-slide imaging: A white paper from the digital pathology association. *J Pathol Inform* 2019;10:9.
- Phukpattaranont P, Boonyaphiphat P. Segmentation of Cancer Cells in Microscopic Images Using Neural Network and Mathematical Morphology. *IEEE 2006 SICE-ICASE International Joint Conference; 2006.* p. 2312-5.
- Saha M, Chakraborty C. Her2Net: A deep framework for semantic segmentation and classification of cell membranes and nuclei in breast cancer evaluation. *IEEE Trans Image Process* 2018;27:2189-200.
- Wang S, Rong R, Yang DM, Fujimoto J, Yan S, Cai L, *et al.* Computational staining of pathology images to study the tumor microenvironment in lung cancer. *Cancer Res* 2020;80:2056-66.
- Prat J, Oncology FCoG. FIGO's staging classification for cancer of the ovary, fallopian tube, and peritoneum: Abridged republication. *J Gynecol Oncol* 2015;26:87-9.



**Supplementary Figure 1:** Correlation heat map of cellular features



**Supplementary Figure 2:** Examples of misclassified cells. (a-c) Exemplifies over-segmentation. (d and e) portions of adipocytes and/or bubbles of fat; (f) heavily pigment laden cells; (g) red blood cells, (h) plasma cell; (i and j) possible noncellular connective tissue fragments or cells with indistinct borders



**Supplementary Figure 3:** Examples of misclassified image patches. (a and b) (Case 5 in Table) harbor cells with indistinct or nearly overlapping cellular borders. The majority of the cells in (c and d) (Case 13) have optically clear nuclear and/or cytoplasmic appearance. Spindle cells aligned in a curvilinear or linear directionality are present in (e) (Case 9) and (f) (Case 22), respectively. Marked variation in cellular shape and size, indicated by the blue arrows (pleomorphism), can be appreciated in (g) (Case 12). Of note, the cells have hyperchromatic nuclei (see c and d for comparison). There is a mixture of cells with a spectrum of nuclear contours and chromatin profile in (h) (Case 5). Cells with prominent nucleoli (orange arrows) are seen in (i and j) (Cases 2 and 5, respectively) (H&E). The image patches in each panel are 512 pixels  $\times$  512 pixels and approximately correspond to  $\times$ 200