

# Triage of 3D pathology data via 2.5D multiple-instance learning to guide pathologist assessments

Gan Gao<sup>1,\*</sup>, Andrew H. Song<sup>2,3,\*</sup>, Fiona Wang<sup>1</sup>, David Brenes<sup>1</sup>, Rui Wang<sup>1</sup>, Sarah S.L. Chow<sup>1</sup>, Kevin W. Bishop<sup>1</sup>, Lawrence D. True<sup>1</sup>, Faisal Mahmood<sup>2,3</sup>, Jonathan T.C. Liu<sup>1</sup>  
<sup>1</sup>University of Washington, <sup>2</sup>Mass General Brigham, <sup>3</sup>Harvard University

gangao@uw.edu, asong@bwh.harvard.edu, jonliu@uw.edu

## Abstract

Accurate patient diagnoses based on human tissue biopsies are hindered by current clinical practice, where pathologists assess only a limited number of thin 2D tissue slices sectioned from 3D volumetric tissue. Recent advances in non-destructive 3D pathology, such as open-top light-sheet microscopy, enable comprehensive imaging of spatially heterogeneous tissue morphologies, offering the feasibility to improve diagnostic determinations. A potential early route towards clinical adoption for 3D pathology is to rely on pathologists for final diagnosis based on viewing familiar 2D H&E-like image sections from the 3D datasets. However, manual examination of the massive 3D pathology datasets is infeasible. To address this, we present CARP3D, a deep learning triage approach that automatically identifies the highest-risk 2D slices within 3D volumetric biopsy, enabling time-efficient review by pathologists. For a given slice in the biopsy, we estimate its risk by performing attention-based aggregation of 2D patches within each slice, followed by pooling of the neighboring slices to compute a context-aware 2.5D risk score. For prostate cancer risk stratification, CARP3D achieves an area under the curve (AUC) of 90.4% for triaging slices, outperforming methods relying on independent analysis of 2D sections (AUC=81.3%). These results suggest that integrating additional depth context enhances the model's discriminative capabilities. In conclusion, CARP3D has the potential to improve pathologist diagnosis via accurate triage of high-risk slices within large-volume 3D pathology datasets.

## 1. Introduction

Disease diagnosis and characterization rely upon the accurate histological analysis of biopsies and surgical specimens by pathologists [19, 33]. In conventional histopathology, only a few thin 2D slices are sectioned from these

tissue specimens for microscopic evaluation. Despite being regarded as the gold standard for medical diagnostics for over a century, conventional histology suffers from severe undersampling of tissue specimens, where less than 1% of a biopsy is examined by a pathologist (4  $\mu\text{m}$  thick sections from 1 mm diameter core biopsy). Furthermore, isolated cross-sectional views of complex 3D tissue structures can be ambiguous and misleading [39, 56]. Recent advances in high-throughput 3D microscopy, along with tissue clearing and fluorescence labeling, now enable non-destructive imaging of large tissue volumes for improved disease characterization, including whole biopsies [5, 17, 18, 24, 33, 38, 49–51, 53, 55]. For example, open-top light-sheet (OTLS) microscopy in conjunction with associated tissue-processing and data-processing methods [3, 5, 20, 21] have shown the ability to generate high-quality 3D pathology datasets at various spatial resolutions with comparable quality to slide-based H&E histology [41, 42].

A number of computational methods have been developed to analyze 3D pathology datasets without human intervention [44, 48, 55], but fully computational analyses require additional large-scale validation studies before they are ready for clinical deployment. To encourage clinical adoption, it is important to show pathologists 2D cross-sectional images that are false colored to mimic the appearance of histology, allowing pathologists to leverage their training and expertise in interpreting 2D hematoxylin and eosin (H&E) images [29, 33, 34, 41]. However, it is infeasible for pathologists to manually analyze large feature-rich 3D pathology datasets containing hundreds or even thousands of 2D slices. Therefore, there is a need for computational triage methods that can efficiently sift through the vast numbers of slices in each 3D pathology dataset and to identify the highest-risk slices for time-efficient pathologist review (**Figure 1**). This approach is a potential low-risk early pathway towards clinical adoption of 3D pathology that keeps pathologists involved in the final diagnosis.

For developing a 3D pathology computational frame-

\*Equal contribution

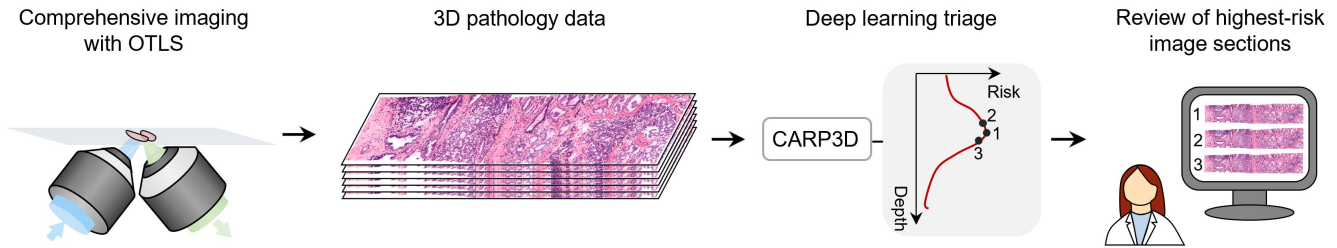


Figure 1. **Workflow with deep-learning-based triage framework for 3D pathology.** Prostate biopsies are comprehensively imaged in 3D with open-top light-sheet (OTLS) microscopy. A deep-learning-based triage method evaluates all 2D slices within 3D pathology datasets and identifies the highest-risk 2D slices for time-efficient pathologist review.

work, it is important to leverage technical advances in 2D computational pathology based on 2D whole-slide images (WSIs), which has witnessed tremendous progress for various clinical outcome prediction tasks such as cancer grading and prognosis, and treatment response predictions [7, 16, 46]. Due to the gigapixel nature of WSIs, these approaches have centered around a multiple instance learning (MIL) paradigm. In typical MIL setups, WSIs are partitioned into a set of smaller patches (i.e., instances), each of which is encoded into a low-dimensional feature vector using models pretrained on natural images [15, 23] or more recently, pretrained on in-domain histopathology images [47, 54]. These patch-level features are then aggregated with aggregation networks into slide-level representations [26, 35, 45, 47]. A direct application of MIL frameworks to each 2D slice within a 3D pathology dataset would allow for risk assessment of each slice. However, the analysis of each 2D slice as an independent image does not take advantage of the added context that exists along the 3rd dimension (depth dimension) of a 3D pathology dataset [33]. For example, a recent work [48] has shown that 3D analyses are superior to 2D (in plane) analyses for a patient risk-stratification task based on 3D pathology. However, that approach was designed to make predictions from 3D volumes based on 3D features, making it ill-suited for pinpointing the most important 2D slices at high granularity, which is essential to guide manual review by a human pathologist.

In this work, we propose **Context-Aware Risk Prediction for 3D pathology (CARP3D)**, a 2.5D MIL framework for risk prediction. CARP3D provides a natural mechanism for incorporating contextual information from neighboring slices, which not only leads to enhanced slice-level predictions but also yields a high-resolution (equal to axial sampling pitch) risk profile along the depth axis for triaging applications. Specifically, slices are patched, featurized, and subsequently aggregated into slice-level feature representations by an intra-slice attention-based network in 2D. An inter-slice pooling module subsequently assigns weights to neighboring images based on their diagnostic importance and integrates neighboring features with respective

weighting factors to assist predictions on the slice of interest. The attention network allows for the capture of fine-grained details within each slice and the inter-slice pooling module emphasizes contextually relevant information while suppressing irrelevant signals for more accurate assessment of a slice of interest. One advantage of our 2.5D approach, as opposed to a full 3D approach, is that it is able to evaluate 3D pathology data at high resolution along the depth axis, and also takes advantage of the emerging family of 2D feature extraction models for 2D pathology images [12, 25, 36, 54], which are lacking in 3D feature extraction.

We apply our method to a cohort of prostate cancer biopsies imaged with OTLS, on the task of discriminating between low-grade (Grade group 1) vs. intermediate- to high-grade prostate cancer (Grade group  $\geq 2$ ), an important clinical task for ensuring that higher-risk patients receive potentially life-saving treatments while low-risk patients are spared serious treatment-related side effects [9]. In terms of predicting risks of given slices, our 2.5D methods outperform the 2D counterparts by a large margin, demonstrating the significance of incorporating 3D contextual information. The codes for this study are publicly available at <https://github.com/alecgao066/CARP3D>.

## 2. Related work

### 2.1. MIL for 2D WSI classification

Since directly analyzing gigapixel WSIs often exceeds the capacities of modern GPUs, most classification tasks follow a MIL framework, also termed weakly-supervised learning. In MIL, the WSI is divided into a set of smaller patches (typically  $256 \times 256$  pixels) from which the low-dimensional features are extracted and aggregated to yield a slide-level feature [47]. Only a single supervisory label is typically provided for the entire WSI. For patch feature encoding, studies have explored encoding patches by transfer learning with a ResNet50 model pretrained on ImageNet [35], as well as augmentations of patch features in low-data regimes [57]. Recently, CTransPath (a

hybrid CNN and Vision Transformer) was trained in a self-supervised learning regime on millions of histopathology images for more generalizable patch features [54]. For patch feature aggregation, different approaches have been explored [8, 30]. Among these methods, attention-based networks are gaining popularity, surpassing rule-based models like max, top-k, or average pooling as a result of superior domain adaptability and data efficiency [13, 14, 30, 32, 35]. Recent aggregation approaches explore either GNN-based aggregation [10, 31] or Transformer self-attention-based aggregations [11, 45, 52] to more explicitly model relationships between patch features.

## 2.2. Computational 3D pathology to support diagnostic determinations

In the realm of 3D pathology, where datasets are massive in size, several AI-assisted pipelines have been explored. Some recent studies have focused on 3D segmentation of diagnostically important tissue morphologies, such as prostate glands and nuclei, to facilitate the extraction of hand-crafted features for patient risk stratification [44, 55]. While intuitive and enabling hypothesis testing/generation, the analysis of a small set of intuitive 3D morphological features cannot fully leverage the complex spatial biomarkers contained within 3D pathology datasets, many of which are opaque to human observers. Recent advances in 3D weakly supervised learning allow for “deep features” to be automatically extracted from tissue volumes for patient prognosis [48]. Such fully-automated end-to-end deep-learning approaches lack explainability and may initially be challenging for clinicians to adopt.

Human-in-the-loop strategies that guide pathologists to review important regions could be a low-risk and attractive strategy for more immediate clinical adoption. Towards this end, a fully supervised 2D deep-learning approach was previously developed to identify high-risk image sections within 3D esophageal data for pathologist review [4], which was based on a limited number of pixel-level annotations painstakingly provided by a single pathologist. In contrast, CARP3D is a weakly-supervised learning pipeline trained on diverse image-level annotations that are easily and quickly generated by a panel of pathologists. Furthermore, it incorporates contextual information in 3D pathology data to achieve better triage performance.

## 2.3. 2.5D analysis for 3D data

2.5D analysis considers 3D volumes as a stack of 2D images and leverages neighboring images to refine predictions, incorporating contextual information with lower computational demands than full 3D models. For example, in medical imaging, several models merge nearby gray-scale images into a single multi-channel image to generate segmentation masks or to detect bounding boxes [2, 40, 59]. Ad-

ditionally, some approaches adopt auxiliary information extracted from nearby images to improve the 2D model outputs, like adjacent prediction results [22] or residuals between neighboring images [58]. Recent works focus on explicitly characterizing context information in 3D datasets with two-stage models, by first conducting 2D analysis for each image and then aggregating features from neighboring images to refine predictions. For such methods, common aggregation approaches include pooling strategies, ensemble methods, or neural networks like RNN [1, 6, 27, 28, 59]. However, the majority of these methods were developed for images considerably smaller than 3D pathology images, thus necessitating a 2.5D method tailored for 3D pathology data, such as CARP3D.

## 3. Methods

We represent each 3D pathology image as a stack of 2D images  $\{\mathbf{X}_i\}_{i=1}^N$ , where  $\mathbf{X}_i \in \mathbb{R}^{W \times H \times C}$  with  $W$ ,  $H$ , and  $C$  denoting width, height, and number of channels respectively, and  $N$  referring to the number of 2D sections in 3D data. For a slice of interest (SOI)  $\mathbf{X}_k$  where  $k \in \{1, \dots, N\}$ , we aim to predict the slice-level label  $y_k$  (Figure 2). For the set of smaller patches from tessellating the SOI and its neighboring slices, we apply a sequence of set aggregation approaches to ultimately construct a single SOI feature for risk prediction. The aim is to harness the extra contextual information provided by the depth dimension: 1) *Intra-slice* attention-based aggregation to construct a slice-level feature and 2) *Inter-slice* aggregation to incorporate neighboring slice-level features to construct a context-aware SOI feature.

We first introduce the patch feature extraction (Section 3.1) and intra-slice attention-based MIL (ABMIL) (Section 3.2). We then describe strategies for inter-slice context aggregation from neighboring slice-level features (Section 3.3), followed by the classification module (Section 3.4) and training/inference steps (Section 3.5).

### 3.1. Patch feature encoding

We denote the set of patches forming an image  $\mathbf{X}_i$  as  $\{\mathbf{x}_i^j\}_{j=1}^J$  where  $\mathbf{x}_i^j \in \mathbb{R}^{w \times h \times C}$ ,  $w$  and  $h$  are lateral dimensions of the patches and  $J$  indicates total number of patches in  $\mathbf{X}_i$ . A pretrained feature encoder  $f_{\text{enc}}$  then extracts a low-dimensional and representative feature  $\mathbf{h}_i^j \in \mathbb{R}^d$  from each patch  $\mathbf{x}_i^j$ , such that  $\mathbf{h}_i^j = f_{\text{enc}}(\mathbf{x}_i^j)$ .

To address the domain differences stemming from utilizing pretrained encoders to encode patches of a different data domain (i.e., pretraining dataset: histology, evaluation dataset: OTLS) [35], we apply a fully-connected layer with ReLU nonlinearity to  $\{\mathbf{h}_i^j\}_{j=1}^J$  to generate more compressed and domain-specific features  $\{\tilde{\mathbf{h}}_i^j\}_{j=1}^J$ , with  $\tilde{\mathbf{h}}_i^j \in \mathbb{R}^{512}$ . Specifically,  $\tilde{\mathbf{h}}_i^j = \text{ReLU}(\mathbf{W}^T \mathbf{h}_i^j + \mathbf{b})$ , where

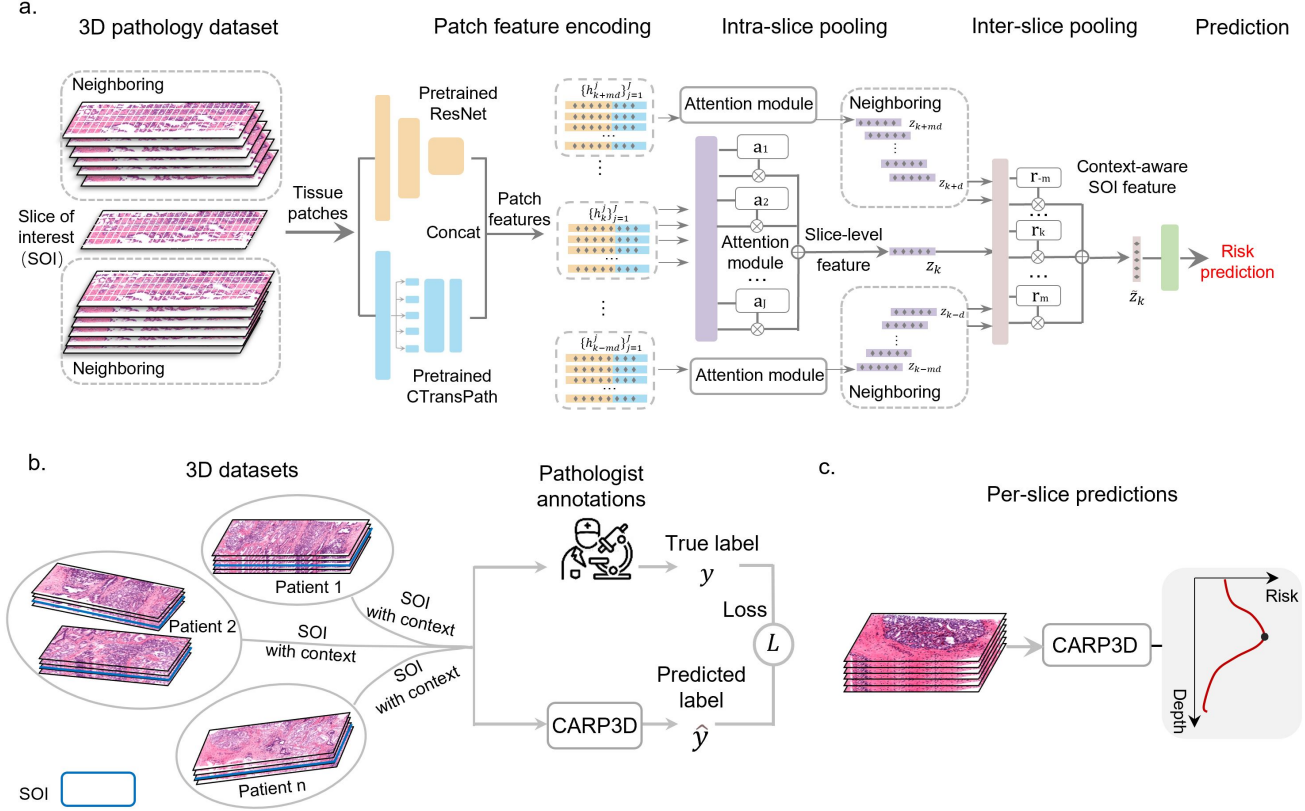


Figure 2. **CARP3D architecture.** a) Patches for a 2D slice of interest (SOI) and its neighboring slices are encoded with pretrained ResNet50 and CTransPath. An intra-slice attention module aggregates patch-level features within each slice into slice-level features. Neighboring slice-level features are aggregated through an inter-slice pooling module to produce a context-aware SOI feature for subsequent risk prediction, formulated as a classification task into high- vs. low-risk categories here. b) During training, slices are selected within each 3D sample of the training set, from which the CARP3D model learns to predict the ground truth labels provided by pathologists. c) Model deployment on 3D pathology data for slice-by-slice risk assessment. The highest-risk slices are selected for pathologist review.

$\mathbf{W} \in \mathbb{R}^{d \times 512}$  and  $\mathbf{b} \in \mathbb{R}^{512}$ .

### 3.2. Intra-slice attention module

The fine-tuned patch features  $\{\tilde{\mathbf{h}}_i^j\}_{j=1}^J$  are aggregated by an attention network [26] into a slice-level feature  $\mathbf{z}_i \in \mathbb{R}^{512}$ , where the attention score  $a_i^j$  computed for each feature  $\tilde{\mathbf{h}}_i^j$  reflects its importance to the final prediction. The attention network is comprised of three sets of parameters  $\mathbf{V} \in \mathbb{R}^{512 \times 256}$ ,  $\mathbf{U} \in \mathbb{R}^{512 \times 256}$ , and  $\mathbf{W} \in \mathbb{R}^{256 \times 1}$  and the attention calculation,

$$a_i^j = \frac{\exp\{\mathbf{W}^T(\tanh(\mathbf{V}^T \tilde{\mathbf{h}}_i^j) \odot \text{sigm}(\mathbf{U}^T \tilde{\mathbf{h}}_i^j))\}}{\sum_{j'=1}^J \exp\{\mathbf{W}^T(\tanh(\mathbf{V}^T \tilde{\mathbf{h}}_i^{j'}) \odot \text{sigm}(\mathbf{U}^T \tilde{\mathbf{h}}_i^{j'}))\}}, \quad (1)$$

where  $\sum_{j=1}^J a_i^j = 1$ ,  $\tanh$  and  $\text{sigm}$  denote hyperbolic tangent and sigmoid function respectively, and  $\odot$  denotes element-wise multiplication operation. The individual patch features are weight-averaged by their correspond-

ing attention scores, resulting in  $\mathbf{z}_i$ ,

$$\mathbf{z}_i = \sum_{j=1}^J a_i^j \tilde{\mathbf{h}}_i^j. \quad (2)$$

While  $\mathbf{z}_i$  alone is sufficient to predict the slice-level label, we next show that leveraging the contextual information from neighboring slices can improve the performance further, which is only possible in 3D pathology.

### 3.3. Inter-slice pooling for 2.5D integration

To leverage contextual information from neighboring images and improve the prediction of the SOI  $\mathbf{X}_k$ , we incorporate  $m$  additional slices above and below  $\mathbf{X}_k$  with the spacing of  $d$  slices in between, i.e.,  $\{\mathbf{X}_{k+id}\}_{i=-m}^m$ . All of these slices are converted to slice-level features following the steps in Section 3.1 and 3.2.

We subsequently apply weighted averaging on the SOI and neighboring slice-level features, where the weights are learnable, to generate a context-aware SOI feature  $\tilde{\mathbf{z}}_k$  for

the downstream prediction task. Specifically, we introduce  $\mathbf{L} \in \mathbb{R}^{512}$  to learn slice-level weights  $\{r_i\}_{i=-m}^m$  such that

$$r_i = \frac{\exp\{\mathbf{L}^T \mathbf{z}_{k+id}\}}{\sum_{i=-m}^m \exp\{\mathbf{L}^T \mathbf{z}_{k+id}\}}. \quad (3)$$

Similar to attention-based weights in intra-slice aggregation, the weights  $\{r_i\}_{i=-m}^m$  emphasize diagnostically-relevant slices. Nevertheless, since this module deals with only a few slice-level features that are already more discernible than patch-level features, we use  $\mathbf{L}$ , which has far fewer parameters than Eq. 1. The feature  $\tilde{\mathbf{z}}_k$  is then calculated as the weighted average of the slice features,

$$\tilde{\mathbf{z}}_k = \sum_{i=-m}^m r_i \mathbf{z}_{k+id}. \quad (4)$$

Here  $r_i$  refers to the weight of  $\mathbf{z}_{k+id}$ , which are summed up to 1.  $\mathbf{L}$  is built to learn to weigh each slice-level feature,  $\mathbf{L} \in \mathbb{R}^{512}$  and its parameters are learned during training.

### 3.4. Classification module

As a final step, the prediction  $\hat{y}_k$  on  $\mathbf{X}_k$  is calculated as follows. Here  $\mathbf{C}$  is the classification layer,  $\mathbf{C} \in \mathbb{R}^{512 \times n}$ , and  $n$  is the number of risk classes.  $n = 2$  for this study.

$$\hat{y}_k = \text{softmax}(\mathbf{C}^T \tilde{\mathbf{z}}_k + b). \quad (5)$$

### 3.5. Training and inference

For the training phase, we select one or two representative slices from each biopsy volume (typically the center slice) as the input and its corresponding pathologists-provided label as the target (**Figure 2(b)**). Since ground truth generation for all slices in each biopsy is infeasible, we extracted representative slices from each 3D dataset to form the training set, promoting the inclusion of images from more diverse patients and biopsies. During the inference phase, the trained model can be deployed across all image sections to generate a predicted risk profile for the volume at the axial sampling pitch ( $1\mu m$ ) of each slice (**Figure 2(c)**). In a clinical context, the image sections with the highest risks would be selected for further evaluation by pathologists.

## 4. Experiments

### 4.1. Data description

We implemented CARP3D for prognostic risk stratification of prostate cancer based on Gleason gradings provided by a panel of 6 board-certified genitourinary pathologists. Simulated core-needle biopsies (roughly  $1 \times 1 \times 15mm$ ) were cut from cancer-enriched regions of archived prostatectomy specimens that were previously formalin-fixed and paraffin-embedded (FFPE). The biopsies were stained with a fluorescent analog of hematoxylin and eosin, optically cleared to

make them transparent, and imaged with OTLS microscopy in 3D to generate 16-bit gray-scale datasets of two fluorescent channels (nuclear stain and cytoplasmic stain) [5, 20]. The sampling pitch of the images is  $\sim 1 \mu m/\text{pixel}$  at 10X-equivalent optical resolution laterally (the typical magnification used for Gleason grading) and  $\sim 4 \mu m$  optical resolution axially (similar to the thickness of standard slide-based histology).

A total of 124 slices were selected from 115 OTLS-imaged prostate biopsies across 54 patients. These images were false-colored to mimic the appearance of H&E histology using a physics-based approach [43]. For prostate cancer, pathologists quantify the aggressiveness of the cancer using the Gleason grading scheme. Grade group 1 (GG=1) is categorized as low-grade prostate cancer, where patients typically opt for active surveillance (i.e. monitoring without aggressive treatment). Grade Group 2 (GG $\geq 2$ ) cancer and above is considered intermediate or high-grade and patients typically receive curative therapy. To enable accurate characterization of higher-grade prostate cancer to provide potentially life-saving treatments, we trained our algorithm to classify slices within 3D prostate datasets as containing low-grade (GG = 1) vs. higher-grade (GG  $\geq 2$ ) prostate cancer (i.e., a binary classification task).

### 4.2. Implementations

2D slices were split into non-overlapping  $256 \times 256$  patches ( $256 \times 256 \mu m$ ). We used the original two-channel 16-bit gray-scale fluorescence images for training CARP3D. Since the cytoplasm channel exhibits relatively uniform signal distributions, we normalized the 16-bit cytoplasm-channel images after cropping out the background (calculated by Otsu thresholding) and outlier signals (99th percentile). For the nuclear channel, which can vary greatly across an image (depending upon the cellularity of various tissue regions), we performed intensity cropping and normalization on individual patches. Finally, we distributed the normalized nuclear and cytoplasmic channels across the first two channels of the three-channel RGB inputs and left the third channel (B channel) empty.

Dual NVIDIA GeForce RTX 3090 GPUs were used for training and inference. We used the Adam optimizer with the constant learning rate of  $2 \times 10^{-4}$  and a batch size of 256 to iteratively minimize the cross-entropy loss for all experiments. For 2.5D analysis, neighboring slices within a range of  $80 \mu m$  above and below each image section were selected – this allows the incorporation of additional information while remaining within the estimated distance of intra-biopsy grading variability [29]. We tested  $m \in \{1, 2, 4, 8\}$ , maintaining  $md = 80$ , and chose best-performing  $m$  for each baseline.

For evaluation, we used the Area Under the ROC Curve (AUC). We also reported the F2 score, which prioritizes

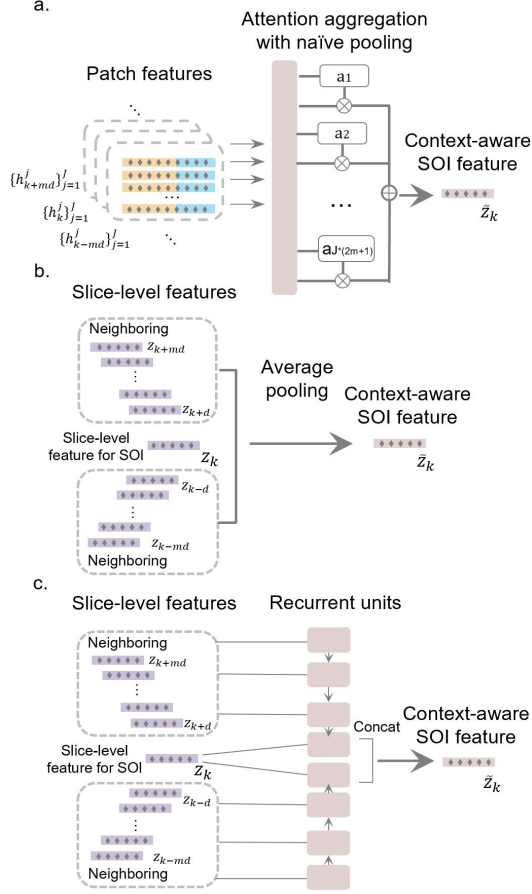


Figure 3. **Baseline architectures for inter-slice pooling.** a) Naive pooling. b) Average pooling. c) RNN-based pooling.

recall vs. precision in comparison to the F1 score which weights them equally. This choice reflects our study’s goal of screening more aggressive prostate cancer patients. The best F2 score is reported by iterating over different thresholds for mapping predicted probabilities to class labels. We employed a leave-one-out cross-validation strategy, where in each fold, slices from one patient are held out to evaluate model performance, and the remaining images are used for training. Predictions were combined across the patients to calculate the cohort-level performance metrics.

### 4.3. Baselines

Our baselines consist of variations to two modules: 1) Patch feature encoding and 2) Inter-slice pooling (Figure 2(a)). All baselines use ABMIL for intra-slice feature aggregation, a lightweight model specifically selected for the data scale.

**Patch feature encoding** We first investigated the most widely-used patch feature encoders, including a ResNet50 [23] pretrained on ImageNet ( $d = 1,024$ ) and a

self-supervised CTransPath (a hybrid CNN / Transformer) pretrained on histopathology images ( $d = 768$ ) [54]. For ResNet50 features, we applied patch augmentations based on random flipping and brightness/contrast jittering to increase the feature diversity (ResNet50<sub>aug</sub>). CTransPath features are not augmented as the self-supervised training of the model makes it invariant to augmentations. Motivated by the success of using multiple features for classification [37], we also tested concatenating the ResNet50 and ResNet50<sub>aug</sub> features with CTransPath features ( $d = 1,792$ ) for more expressive representations.

**Inter-slice pooling** We explored various approaches for inter-slice context aggregation (Figure 3).

- 1) **No pooling** Only patch features from  $\mathbf{X}_k$  are aggregated, and therefore no contextual information is used.
- 2) **Naive pooling** All patch features from  $\{\mathbf{X}_{k+id}\}_{i=-m}^m$  are aggregated by a single attention module to construct  $\tilde{\mathbf{z}}_k$ , disregarding slice identity.
- 3) **Average pooling** Patch features are first aggregated to slice-level features within each image. The neighboring slice-level features are averaged with SOI feature to form  $\tilde{\mathbf{z}}_k$ , where  $\tilde{\mathbf{z}}_k = \frac{1}{(2m+1)} \sum_{i=-m}^m \mathbf{z}_{k+id}$ .
- 4) **RNN-based pooling** Since average pooling is insensitive to the order of images, we tested RNN-based pooling to incorporate sequential information. Recurrent units aggregate image sequences  $\{\mathbf{X}_{k+id}\}_{i=-m}^m$  in a bi-directional manner, integrating slice-level features into the hidden state  $\mathbf{hid}$  as contextual information,

$$\begin{aligned} \mathbf{hid}_{i,1} &= \tanh(\mathbf{W}_n \mathbf{z}_{k+id} + \mathbf{W}_h \mathbf{hid}_{i+1,1}) \\ \mathbf{hid}_{i,-1} &= \tanh(\mathbf{W}_n \mathbf{z}_{k+id} + \mathbf{W}_h \mathbf{hid}_{i-1,-1}), \end{aligned} \quad (6)$$

where  $\mathbf{W}_n, \mathbf{W}_h \in \mathbb{R}^{512 \times 512}$  are weight matrices and +1 or -1 denotes the direction of integration. We concatenate the hidden state features at level  $k$  to form  $\tilde{\mathbf{z}}_k$ , where  $\tilde{\mathbf{z}}_k = \mathbf{hid}_{k,1} \oplus \mathbf{hid}_{k,-1}$ .

## 5. Results

### 5.1. Patch feature encoders

The results for feature encoder ablation can be found in Table 1. We observe that the concatenation of augmented ResNet50 and CTransPath features outperformed all other strategies using 2D attention-based MIL. We observe that ResNet50 features and augmented ResNet50 features achieved a similar AUC, but augmentation helped slightly with identifying the positive class (better F2 score). CTransPath features improved AUC by a significant margin. However, the F2 score is inferior to experiments using original and augmented ResNet50 features, which indicate the improved AUC mainly comes from correctly predicting GG=1 images, while many GG≥2 images are still confused with the low-grade class (i.e. poor recall or diagnostic sensitivity). Concatenation of ResNet50 and CTransPath fea-

Encoders	AUC ( $\uparrow$ )	F2 ( $\uparrow$ )
ResNet50	75.7% (68.2%-82.6%)	87.5% (84.3%-90.9%)
ResNet50 <sub>aug</sub>	75.6% (67.9%-82.3%)	87.9% (84.5%-90.9%)
CTransPath	80.1% (73.3%-86.4%)	86.7% (83.6%-90.3%)
ResNet50 $\oplus$ CTransPath	79.7% (73%-86.2%)	87.0% (83.8%-90.5%)
ResNet50 <sub>aug</sub> $\oplus$ CTransPath	<b>81.3%</b> <b>(75.2%-87.3%)</b>	<b>88.5%</b> <b>(85.5%-91.7%)</b>

Table 1. **Results for different encoders.** Best performances are in **bold**. The 95% confidence interval (95% CI) is calculated based on bootstrapping. Concatenation is denoted by  $\oplus$ .

tures results in a better F2 score than CTransPath, but at the expense of slightly lower AUC. Applying patch augmentation on ResNet50 encoders and concatenating the features with CTransPath features achieved the best AUC and F2 score compared with all other experiments. Therefore, we used these features for subsequent analyses. We leave utilizing the patch features from very recent foundation models as future work [12, 36].

Aggregation	AUC ( $\uparrow$ )	F2 ( $\uparrow$ )
None	81.3% (75.2%-87.3%)	88.5% (85.5%-91.7%)
Naive	87.2% (81.5%-82.2%)	88.2% (85.7%-92.4%)
Average	89.8% (85.1%-94%)	90.1% (87.5%-93.6%)
RNN	89.3% (84.2%-93.6%)	90.5% (87%-93.6%)
Weighted Average	<b>90.1%</b> <b>(85.3%-94.3%)</b>	<b>92.4%</b> <b>(90.2%-95.1%)</b>

Table 2. **Results for inter-slice aggregation approaches.** All methods use the concatenation of ResNet50<sub>aug</sub> and CTransPath features. Best performances are in **bold**. The 95% confidence interval (95% CI) is calculated based on bootstrapping.

## 5.2. Performance comparison between 2.5D vs 2D

The comparisons between 2D and 2.5D approaches can be found in Table 2. We observe that 2.5D methods using contextual information improved the classification of low-grade vs. higher-grade cancer over 2D analysis. We considered attention-based MIL on independent 2D images with no inter-slice aggregation as the baseline and explored different inter-slice context aggregation strategies. The *naive pooling* baseline, where a single attention-based network aggregates

patches from the SOI and neighboring slices, results in notable improvement in AUC, suggesting that additional spatial context provided by the 3rd dimension enhances the discriminative capabilities of the model. However, there was a drop in F2 (poorer recall or diagnostic sensitivity), potentially due to the increased number of patches to attend to, making it more challenging to correctly recognize positive patches containing small foci of aggressive cancer.

To better utilize the depth information within each biopsy, we further analyzed strategies for a sequence of aggregations, first within each slice (ABMIL) followed by across slices. For inter-slice aggregation, we observe that simple averaging of slice features (*average pooling*) achieves further improvements in both AUC and F2. Interestingly, the *RNN-based pooling*, designed to better incorporate sequential information from neighboring images, performs on par with the average pooling. We conjecture that the introduction of more trainable parameters with RNN likely leads to overfitting. Finally, since one limitation of average pooling is that positive images may still be diluted by neighboring images, especially when aggressive cancer constitutes only a small proportion, we designed *weighted averaging* to emphasize slice features that are more diagnostically important. We observe that weighted averaging achieved a significant improvement in AUC and especially in F2. We conclude that a sequence of aggregation based on weighted averaging employed by CARP3D can best utilize the 3D dataset for clinical prediction.

## 5.3. Interpretability

To gain a better understanding of the CARP3D predictions, we applied principal component analysis (PCA) to the context-aware SOI features in **Figure 4(a)**.  $GG \geq 2$  slices are visually separable from  $GG = 1$  slices in the PC space, suggesting that our model can discriminate between two classes. By visualizing heatmaps of attention scores overlaid on top of false-colored H&E-like images, we can identify regions important for rendering predictions. For example, regions in **Figure 4(b)** are predicted as highly likely to contain higher-grade prostate cancer. The high-attention regions are roughly aligned with fused glands, which are commonly associated with higher-grade prostate cancer. Regions in **Figure 4(c)** are predicted as low-grade prostate cancer, where the highly attended regions correspond to well-formed benign glands and lymphocytes.

## 5.4. Triage on a 3D pathology dataset

We demonstrate the slice-by-slice inference of the trained CARP3D on an example 3D dataset (**Figure 5(a)**). Specifically, we generate a depth profile of predicted context-aware risk scores, with the risk defined as the predicted probability of a given slice containing higher-grade prostate cancer. To validate our findings, a board-certified pathologist reviewed

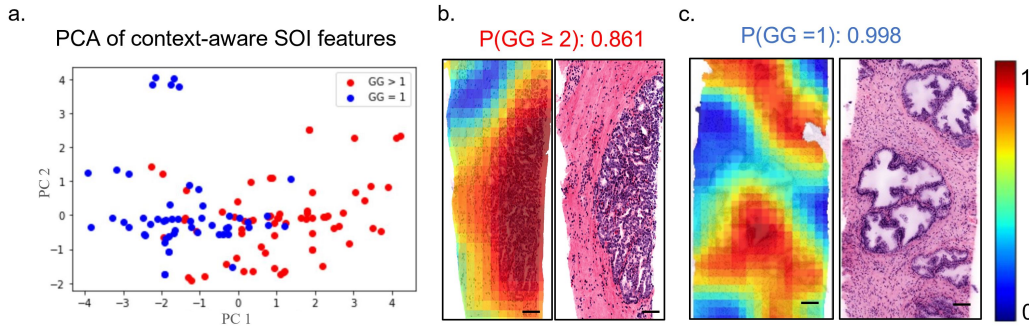


Figure 4. **Visualization of SOI features and interpretable attention heatmaps.** a) PCA of context-aware SOI features. b) and c) show examples of attention heatmaps with corresponding false-colored images. The scale bar is  $100 \mu\text{m}$ . Color bar indicates attention scores.

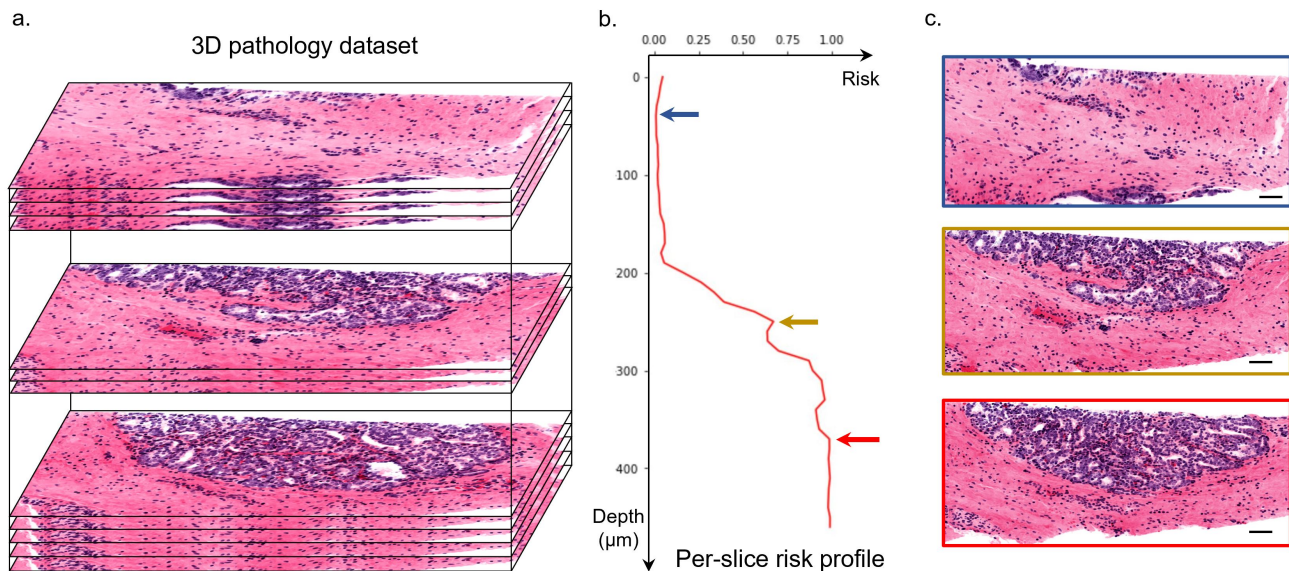


Figure 5. **CARP3D triage on an example 3D pathology dataset.** a) An example 3D pathology dataset. b) Per-slice risk profile, predicted by CARP3D, for higher-grade prostate cancer. c) Images at arrow positions are reviewed by a board-certified pathologist, showing that human evaluation on select slices broadly aligns with the risk profile. The scale bar is  $100 \mu\text{m}$ .

images at the depths indicated by arrows (**Figure 5(b)**). At the blue arrow, only well-formed benign glands are seen (Grade Group 1), albeit with only a small number of glands visible at this depth. Fused glands (associated with higher-grade prostate cancer) are present in the slice at the brown arrow, but benign glands are more prevalent (Grade Group 2). The fused glands become the predominant morphology at the red arrow, suggesting more aggressive cancer (Grade Group 3). In summary, the pathologist evaluation of select slices broadly aligns with the risk profile generated by our model. In real-world clinical practice, the highest risk slice at the red arrow as in **Figure 5(c)** would be prioritized for pathologist review. A large-scale clinical study will be performed in the future to demonstrate the ability of AI-triaged 3D pathology (enabled by CARP3D) to improve the detec-

tion of higher-grade prostate cancer in comparison to standard 2D histopathology.

## 6. Conclusion

We present CARP3D, a 2.5D multiple instance learning framework to triage the highest-risk slices within 3D pathology datasets to facilitate pathologist review. Our work leverages contextual information in 3D pathology data to enhance the predicting accuracy of each slice. CARP3D could potentially accelerate the clinical adoption of 3D pathology by improving pathologists' diagnostic accuracy via increased tissue sampling and context-aware triage. Future work includes curation of more 3D pathology data across organs for large-scale clinical validation in comparison to standard 2D histopathology.



## References

- [1] Vibhu Agrawal, Jayaram Udupa, Yubing Tong, and Drew Torigian. BRR-Net: A tandem architectural cnn-rnn for automatic body region localization in ct images. *Medical Physics*, 47(10):5020–5031, 2020. [3](#)
- [2] Arman Avesta, Sajid Hossain, MingDe Lin, Mariam Aboian, Harlan M Krumholz, and Sanjay Aneja. Comparing 3D, 2.5 D, and 2D approaches to brain image auto-segmentation. *Bioengineering*, 10(2):181, 2023. [3](#)
- [3] Lindsey A Barner, Adam K Glaser, Hongyi Huang, Lawrence D True, and Jonathan TC Liu. Multi-resolution open-top light-sheet microscopy to enable efficient 3D pathology workflows. *Biomedical Optics Express*, 11(11):6605–6619, 2020. [1](#)
- [4] Lindsey A Erion Barner, Gan Gao, Deepti M Reddi, Lydia Lan, Wynn Burke, Faisal Mahmood, William M Grady, and Jonathan TC Liu. AI-triaged 3D pathology to improve detection of esophageal neoplasia while reducing pathologist workloads. *Modern Pathology*, 36(12):100322, 2023. [3](#)
- [5] Kevin W Bishop, Lindsey A Erion Barner, Qinghua Han, Elena Baraznenok, Lydia Lan, Chetan Poudel, Gan Gao, Robert B Serafin, Sarah SL Chow, Adam K Glaser, et al. An end-to-end workflow for nondestructive 3D pathology. *Nature Protocols*, pages 1–27, 2024. [1](#), [5](#)
- [6] David Brenes, Mila P. Salcedo, Jackson B. Coole, Yajur Maker, Alex Kortum, Richard A. Schwarz, Jennifer Carns, Imran S. Vohra, Júlio C. Possati-Resende, Márcio Antoniazzi, Bruno de Oliveira Fonseca, Karen C. Borba Souza, Iara V. Vidigal Santana, Flávia Fazzio Barbin, Regis Kreitchmann, Nirmala Ramanujam, Kathleen M. Schmeler, and Rebecca Richards-Kortum. Multiscale optical imaging fusion for cervical precancer diagnosis: Integrating widefield colposcopy and high-resolution endomicroscopy. *IEEE Transactions on Biomedical Engineering*, pages 1–10, 2024. [3](#)
- [7] Wouter Bulten, Kimmo Kartasalo, Po-Hsuan Cameron Chen, Peter Ström, Hans Pinckaers, Kunal Nagpal, Yuannan Cai, David F Steiner, Hester Van Boven, Robert Vink, et al. Artificial intelligence for diagnosis and gleason grading of prostate cancer: the panda challenge. *Nature medicine*, 28(1):154–163, 2022. [2](#)
- [8] Iain Carmichael, Andrew H Song, Richard J Chen, Drew FK Williamson, Tiffany Y Chen, and Faisal Mahmood. Incorporating intratumoral heterogeneity into weakly-supervised deep learning models via variance pooling. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 387–397. Springer, 2022. [3](#)
- [9] Peter R Carroll, J Kellogg Parsons, Gerald Andriole, Robert R Bahnson, Erik P Castle, William J Catalona, Douglas M Dahl, John W Davis, Jonathan I Epstein, Ruth B Etzioni, et al. Prostate cancer early detection, version 2.2016: Featured updates to the nccn guidelines. *JNCCN Journal of the National Comprehensive Cancer Network*, 14(5):509–519, 2016. [2](#)
- [10] Richard J Chen, Ming Y Lu, Muhammad Shaban, Chengkuan Chen, Tiffany Y Chen, Drew FK Williamson, and Faisal Mahmood. Whole slide images are 2d point clouds: Context-aware survival prediction using patch-based graph convolutional networks. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24*, pages 339–349. Springer, 2021. [3](#)
- [11] Richard J Chen, Chengkuan Chen, Yicong Li, Tiffany Y Chen, Andrew D Trister, Rahul G Krishnan, and Faisal Mahmood. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16144–16155, 2022. [3](#)
- [12] Richard J Chen, Tong Ding, Ming Y Lu, Drew FK Williamson, Guillaume Jaume, Andrew H Song, Bowen Chen, Andrew Zhang, Daniel Shao, Muhammad Shaban, et al. Towards a general-purpose foundation model for computational pathology. *Nature Medicine*, pages 1–13, 2024. [2](#), [7](#)
- [13] Nicolas Coudray, Paolo Santiago Ocampo, Theodore Sakellaropoulos, Navneet Narula, Matija Snuderl, David Fenyö, Andre L Moreira, Narges Razavian, and Aristotelis Tsirigós. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nature medicine*, 24(10):1559–1567, 2018. [3](#)
- [14] Pierre Courtiol, Charles Maussion, Matahi Moarii, Elodie Pronier, Samuel Pilcer, Meriem Sefta, Pierre Manceron, Sylvain Toldo, Mikhail Zaslavskiy, Nolwenn Le Stang, et al. Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nature medicine*, 25(10):1519–1525, 2019. [3](#)
- [15] 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*, pages 248–255. Ieee, 2009. [2](#)
- [16] Saman Farahmand, Aileen I Fernandez, Fahad Shabbir Ahmed, David L Rimm, Jeffrey H Chuang, Emily Reisenbichler, and Kourosh Zarringhalam. Deep learning trained on hematoxylin and eosin tumor region of interest predicts HER2 status and trastuzumab treatment response in her2+ breast cancer. *Modern Pathology*, 35(1):44–51, 2022. [2](#)
- [17] André Forjaz, Eduarda Vaz, Valentina Matos Romero, Saurabh Joshi, Alicia M. Braxton, Ann C. Jiang, Kohei Fujikura, Toby Cornish, Seung-Mo Hong, Ralph H. Hruban, Pei-Hsun Wu, Laura D. Wood, Ashley L. Kiemen, and Dennis Wirtz. Three-dimensional assessments are necessary to determine the true spatial tissue composition of diseased tissues. *bioRxiv*, 2023. [1](#)
- [18] Jasper Frohn, Diana Pinkert-Leetsch, Jeannine Missbach-Güntner, Marius Reichardt, Markus Osterhoff, Frauke Alves, and Tim Salditt. 3D virtual histology of human pancreatic tissue by multiscale phase-contrast X-ray tomography. *Journal of Synchrotron Radiation*, 27(6):1707–1719, 2020. [1](#)
- [19] Gan Gao, Dominic Miyasato, Lindsey A Barner, Robert Serafin, Kevin W Bishop, Weisi Xie, Adam K Glaser, Eben L Rosenthal, Lawrence D True, and Jonathan TC Liu. Comprehensive surface histology of fresh resection margins with rapid Open-Top Light-Sheet (OTLS) microscopy. *IEEE Transactions on Biomedical Engineering*, 2023. [1](#)

- [20] Adam K Glaser, Nicholas P Reder, Ye Chen, Chengbo Yin, Linpeng Wei, Soyoun Kang, Lindsey A Barner, Weisi Xie, Erin F McCarty, Chenyi Mao, et al. Multi-immersion open-top light-sheet microscope for high-throughput imaging of cleared tissues. *Nature communications*, 10(1):2781, 2019. 1, 5
- [21] Adam K Glaser, Kevin W Bishop, Lindsey A Barner, Et-suo A Susaki, Shimpei I Kubota, Gan Gao, Robert B Serafin, Pooja Balaram, Emily Turschak, Philip R Nicovich, et al. A hybrid open-top light-sheet microscope for versatile multi-scale imaging of cleared tissues. *Nature methods*, 19(5): 613–619, 2022. 1
- [22] Yesenia Gonzalez, Chenyang Shen, Hyunuk Jung, Dan Nguyen, Steve B Jiang, Kevin Albuquerque, and Xun Jia. Semi-automatic sigmoid colon segmentation in ct for radiation therapy treatment planning via an iterative 2.5-d deep learning approach. *Medical image analysis*, 68:101896, 2021. 3
- [23] 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*, pages 770–778, 2016. 2, 6
- [24] Seung-Mo Hong, DongJun Jung, Ashley Kiemen, Matthias M. Gaida, Tadashi Yoshizawa, Alicia M. Braxton, Michaël Noë, Gemma Lionheart, Kiyoko Oshima, Elizabeth D. Thompson, Richard Burkhart, Pei-Hsun Wu, Denis Wirtz, Ralph H. Hruban, and Laura D. Wood. Three-dimensional visualization of cleared human pancreas cancer reveals that sustained epithelial-to-mesenchymal transition is not required for venous invasion. *Modern Pathology*, 33(4):639–647, 2020. 1
- [25] Zhi Huang, Federico Bianchi, Mert Yuksekgonul, Thomas J Montine, and James Zou. A visual–language foundation model for pathology image analysis using medical twitter. *Nature medicine*, 29(9):2307–2316, 2023. 2
- [26] Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International conference on machine learning*, pages 2127–2136. PMLR, 2018. 2, 4
- [27] Alexander Ke, Shih-Cheng Huang, Chloe P O’Connell, Michal Klimont, Serena Yeung, and Pranav Rajpurkar. Video pretraining advances 3d deep learning on chest ct tasks. In *Medical Imaging with Deep Learning*, pages 758–774. PMLR, 2024. 3
- [28] Titinunt Kitrungrotsakul, Xian-Hau Han, Yutaro Iwamoto, Satoko Takemoto, Hideo Yokota, Sari Ipponjima, Tomomi Nemoto, Wei Xiong, and Yen-Wei Chen. A cascade of 2.5 d cnn and bidirectional clstm network for mitotic cell detection in 4d microscopy image. *IEEE/ACM transactions on computational biology and bioinformatics*, 18(2):396–404, 2019. 3
- [29] Can Koyuncu, Andrew Janowczyk, Xavier Farre, Tilak Pathak, Tuomas Mirtti, Pedro L Fernandez, Laura Pons, Nicholas P Reder, Robert Serafin, Sarah SL Chow, et al. Visual assessment of 2-dimensional levels within 3-dimensional pathology data sets of prostate needle biopsies reveals substantial spatial heterogeneity. *Laboratory Investigation*, 103(12):100265, 2023. 1, 5
- [30] Narmin Ghaffari Laleh, Hannah Sophie Muti, Chiara Maria Lavinia Loeffler, Amelie Echle, Oliver Lester Sal-danha, Faisal Mahmood, Ming Y Lu, Christian Trautwein, Rupert Langer, Bastian Dislich, et al. Benchmarking weakly-supervised deep learning pipelines for whole slide classification in computational pathology. *Medical image analysis*, 79:102474, 2022. 3
- [31] Yongju Lee, Jeong Hwan Park, Sohee Oh, Kyoungseob Shin, Jiyu Sun, Minsun Jung, Cheol Lee, Hyojin Kim, Jin-Haeng Chung, Kyung Chul Moon, et al. Derivation of prognostic contextual histopathological features from whole-slide images of tumours via graph deep learning. *Nature Biomedical Engineering*, pages 1–15, 2022. 3
- [32] Jana Lipkova, Tiffany Y Chen, Ming Y Lu, Richard J Chen, Maha Shady, Mane Williams, Jingwen Wang, Zahra Noor, Richard N Mitchell, Mehmet Turan, et al. Deep learning-enabled assessment of cardiac allograft rejection from endomyocardial biopsies. *Nature medicine*, 28(3):575–582, 2022. 3
- [33] Jonathan TC Liu, Adam K Glaser, Kaustav Bera, Lawrence D True, Nicholas P Reder, Kevin W Eliceiri, and Anant Madabhushi. Harnessing non-destructive 3D pathology. *Nature biomedical engineering*, 5(3):203–218, 2021. 1, 2
- [34] Jonathan TC Liu, Sarah SL Chow, Richard Colling, Michelle R Downes, Xavier Farré, Peter Humphrey, Andrew Janowczyk, Tuomas Mirtti, Clare Verrill, Inti Zlobec, et al. Engineering the future of 3D pathology. *The Journal of Pathology: Clinical Research*, 10(1):e347, 2024. 1
- [35] Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Richard J Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, 5(6):555–570, 2021. 2, 3
- [36] Ming Y Lu, Bowen Chen, Drew FK Williamson, Richard J Chen, Ivy Liang, Tong Ding, Guillaume Jaume, Igor Odintsov, Long Phi Le, Georg Gerber, et al. A visual-language foundation model for computational pathology. *Nature Medicine*, pages 1–12, 2024. 2, 7
- [37] Long D Nguyen, Dongyun Lin, Zhiping Lin, and Jiuwen Cao. Deep cnns for microscopic image classification by exploiting transfer learning and feature concatenation. In *2018 IEEE international symposium on circuits and systems (IS-CAS)*, pages 1–5. IEEE, 2018. 6
- [38] Eben Olson, Michael J Levene, and Richard Torres. Multiphoton microscopy with clearing for three dimensional histology of kidney biopsies. *Biomedical optics express*, 7(8): 3089–3096, 2016. 1
- [39] Stephen M Olson, Mohammad Hussaini, and James S Lewis Jr. Frozen section analysis of margins for head and neck tumor resections: reduction of sampling errors with a third histologic level. *Modern Pathology*, 24(5):665–670, 2011. 1
- [40] Sanson TS Poon, Fahmy WF Hanna, François Lemarchand, Cherian George, Alexander Clark, Simon Lea, Charlie Coleman, and Giuseppe Sollazzo. Detecting adrenal lesions on 3d ct scans using a 2.5 d deep learning model. *medRxiv*, pages 2023–02, 2023. 3

- [41] Deepti M Reddi, Lindsey A Barner, Wynn Burke, Gan Gao, William M Grady, and Jonathan TC Liu. Nondestructive 3D pathology image atlas of Barrett esophagus with open-top light-sheet microscopy. *Archives of Pathology & Laboratory Medicine*, 147(10):1164–1171, 2023. [1](#)
- [42] Nicholas P Reder, Adam K Glaser, Erin F McCarty, Ye Chen, Lawrence D True, and Jonathan TC Liu. Open-top light-sheet microscopy image atlas of prostate core needle biopsies. *Archives of pathology & laboratory medicine*, 143(9):1069–1075, 2019. [1](#)
- [43] Robert Serafin, Weisi Xie, Adam K Glaser, and Jonathan TC Liu. Falsecolor-python: a rapid intensity-leveling and digital-staining package for fluorescence-based slide-free digital pathology. *Plos one*, 15(10):e0233198, 2020. [5](#)
- [44] Robert Serafin, Can Koyuncu, Weisi Xie, Hongyi Huang, Adam K Glaser, Nicholas P Reder, Andrew Janowczyk, Lawrence D True, Anant Madabhushi, and Jonathan TC Liu. Nondestructive 3D pathology with analysis of nuclear features for prostate cancer risk assessment. *The Journal of Pathology*, 260(4):390–401, 2023. [1](#), [3](#)
- [45] Zhuchen Shao, Hao Bian, Yang Chen, Yifeng Wang, Jian Zhang, Xiangyang Ji, et al. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. *Advances in neural information processing systems*, 34:2136–2147, 2021. [2](#), [3](#)
- [46] Ole-Johan Skrede, Sepp De Raedt, Andreas Kleppe, Tarjei S Hveem, Knut Liestøl, John Maddison, Hanne A Askautrud, Manohar Pradhan, John Arne Nesheim, Fritz Albrechtsen, et al. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. *The Lancet*, 395(10221):350–360, 2020. [2](#)
- [47] Andrew H Song, Guillaume Jaume, Drew FK Williamson, Ming Y Lu, Anurag Vaidya, Tiffany R Miller, and Faisal Mahmood. Artificial intelligence for digital and computational pathology. *Nature Reviews Bioengineering*, 1(12):930–949, 2023. [2](#)
- [48] Andrew H Song, Mane Williams, Drew FK Williamson, Sarah SL Chow, Guillaume Jaume, Gan Gao, Andrew Zhang, Bowen Chen, Alexander S Baras, Robert Serafin, Richard Colling, Michelle R Downes, Xavier Farre, Peter Humphrey, Clare Verrill, Lawrence D True, Anil V Parwani, Jonathan TC Liu, and Faisal Mahmood. Analysis of 3D pathology samples using weakly supervised AI. *Cell*, 2024. [1](#), [2](#), [3](#)
- [49] Nobuyuki Tanaka, Shigeaki Kanatani, Raju Tomer, Cecilia Sahlgren, Pauliina Kronqvist, Dagmara Kaczynska, Lauri Louhivuori, Lorand Kis, Claes Lindh, Przemysław Mitura, et al. Whole-tissue biopsy phenotyping of three-dimensional tumours reveals patterns of cancer heterogeneity. *Nature Biomedical Engineering*, 1(10):796–806, 2017. [1](#)
- [50] Rong Tang, Julliette M Buckley, Leopoldo Fernandez, Suzanne Coopey, Owen Aftreth, James Michaelson, Mansi Saksena, Lan Lei, Michelle Specht, Michele Gadd, et al. Micro-computed tomography (Micro-CT): a novel approach for intraoperative breast cancer specimen imaging. *Breast cancer research and treatment*, 139:311–316, 2013.
- [51] Martin E van Royen, Esther I Verhoef, Charlotte F Kweldam, Wiggert A van Cappellen, Gert-Jan Kremers, Adrian B Houtsmuller, and Geert JLH van Leenders. Three-dimensional microscopic analysis of clinical prostate specimens. *Histopathology*, 69(6):985–992, 2016. [1](#)
- [52] Sophia J Wagner, Daniel Reisenbüchler, Nicholas P West, Jan Moritz Niehues, Jiefu Zhu, Sebastian Foersch, Gregory Patrick Veldhuizen, Philip Quirke, Heike I Grabsch, Piet A van den Brandt, et al. Transformer-based biomarker prediction from colorectal cancer histology: A large-scale multicentric study. *Cancer Cell*, 41(9):1650–1661, 2023. [3](#)
- [53] Linghua Wang, Mingyao Li, and Tae Hyun Hwang. The 3d revolution in cancer discovery. *Cancer Discovery*, 14(4):625–629, 2024. [1](#)
- [54] 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. [2](#), [3](#), [6](#)
- [55] Weisi Xie, Nicholas P Reder, Can Koyuncu, Patrick Leo, Sarah Hawley, Hongyi Huang, Chenyi Mao, Nadia Postupna, Soyoung Kang, Robert Serafin, et al. Prostate cancer risk stratification via nondestructive 3D pathology with deep learning–assisted gland analysis. *Cancer research*, 82(2):334–345, 2022. [1](#), [3](#)
- [56] Achla Bharti Yadav, Mala Kamboj, Anjali Narwal, and Anju Devi. Diagnostic efficacy of deeper sections in routine oral histopathology practice: a retrospective study. *Journal of Dentistry*, 19(1):63, 2018. [1](#)
- [57] Imaad Zaffar, Guillaume Jaume, Nasir Rajpoot, and Faisal Mahmood. Embedding space augmentation for weakly supervised learning in whole-slide images. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, pages 1–4. IEEE, 2023. [2](#)
- [58] Jianpeng Zhang, Yutong Xie, Yan Wang, and Yong Xia. Inter-slice context residual learning for 3d medical image segmentation. *IEEE Transactions on Medical Imaging*, 40(2):661–672, 2020. [3](#)
- [59] Yichi Zhang, Qingcheng Liao, Le Ding, and Jicong Zhang. Bridging 2D and 3D segmentation networks for computation-efficient volumetric medical image segmentation: An empirical study of 2.5 d solutions. *Computerized Medical Imaging and Graphics*, 99:102088, 2022. [3](#)