



Joint Prediction of Response to Therapy, Molecular Traits, and Spatial Organisation in Colorectal Cancer Biopsies

Ruby Wood^{1,3}, Enric Domingo⁴, Korsuk Sirinukunwattana^{1,3},
Maxime W. Lafarge⁵, Viktor H. Koelzer⁵, Timothy S. Maughan⁴,
and Jens Rittscher^{1,2,3}✉

¹ Department of Engineering Science, University of Oxford, Oxford, UK
{ruby.wood,jens.rittscher}@eng.ox.ac.uk

² Nuffield Department of Medicine, University of Oxford, Oxford, UK

³ Big Data Institute, University of Oxford, Li Ka Shing Centre for Health
Information and Discovery, Oxford, UK

⁴ Department of Oncology, University of Oxford, Oxford, UK

⁵ Computational and Translational Pathology Group, Department of Pathology
and Molecular Pathology, University Hospital and University of Zürich,
Zürich 8091, Switzerland

Abstract. Existing methods for interpretability of model predictions are largely based on technical insights and are not linked to clinical context. We use the question of predicting response to radiotherapy in colorectal cancer patients as an exemplar for developing prediction models that do provide such contextual information and therefore can effectively support clinical decision making. There is a growing body of evidence that about 30% of colorectal cancer patients do not respond to radiotherapy and will need alternative treatment. The consensus molecular subtypes for colorectal cancer (CMS) provide one such approach to categorising patients based on their disease biology. Here we select the CMS4 subtype as a proxy for stromal infiltration. By jointly predicting a patient's response to radiotherapy, the presence of CMS4, and the epithelial tissue map from morphological features extracted from standard H&E slides we provide a comprehensive clinically relevant assessment of a biopsy. A graph neural network is trained to achieve this joint prediction task, which subsequently provides novel interpretability maps to aid clinicians in their cancer treatment decision making process. Our model is trained and validated on two private rectal cancer datasets.

Keywords: Histology · Graphs · Colorectal Cancer · Response to Therapy · Interpretability

Supported by Cancer Research UK.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43904-9_73.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14224, pp. 758–767, 2023.
https://doi.org/10.1007/978-3-031-43904-9_73

1 Introduction

In the UK, approximately 11,500 patients are diagnosed with rectal cancer each year [19]. A common form of treatment for such patients is neoadjuvant therapy, including chemotherapy and radiotherapy, which can be given to patients with locally advanced rectal cancer to shrink the tumour prior to surgery. Recent evidence suggests that 10–20% of patients will have a complete pathological response to neoadjuvant therapy and can therefore avoid surgery altogether [2, 5]. However, one third of patients do not benefit from radiotherapy treatment prior to surgery [8], hence it is important to determine how a patient will respond to radiotherapy with a personalized approach in order to avoid overtreatment.

Histology-based digital biomarkers enable the possibility to predict a patient’s response to therapy. The consensus molecular subtypes (CMS) classification system derived from gene expressions [9] has been developed to provide biological insight into metastatic colorectal cancer. It has been shown that these four CMS classes can be predicted directly from the standard haematoxylin and eosin (H&E) stained slide images using deep learning [18]. Various studies have investigated the link between CMS and patient outcomes, suggesting that patients with tumour classified as CMS4, which features stromal invasion [9] and shows significantly higher stroma content [15], have worse survival rates compared to the other CMS classes [5]. Increased stromal content has independently been shown to be a predictor for increased risk of recurrence in early rectal cancer [10], and tumour immune infiltrate evaluated with Immunoscore is a useful prognostic marker [3]. The spatial organisation of the cancerous tissue has been identified as a biomarker for aggressiveness or recurrence [12], and Qi et al. [15] found that the features they developed representing spatial organisation reflected characteristics of the four CMS classes. Interactions between the epithelial tissue (cellular tissue lining) and other prevalent tissue types in the tumour microenvironment are also indicators of prognosis [15], since progression of colorectal cancer is dependent on both the epithelial and stromal tissues [20]. Other work has looked at predicting chemoradiotherapy response in rectal cancer patients from H&E images using different approaches, but without providing contextual interpretations [19, 22].

As opposed to predicting response to radiotherapy alone, we aim to analyse this prediction in the context of the overall tissue architecture and the tumour biology as captured by CMS. Input to our model is a standard H&E Whole Slide Image (WSI) which is split into smaller patches to overcome the memory limitations of existing GPUs. To achieve our goal we need to capture the heterogeneity at the slide level, which is why applying full or semi-supervised approaches on individual tiles followed by a slide aggregation method is not suitable. Instead, we build on recent graph neural network (GNN) approaches that allow us to model the entire WSI as a graph. As local cell communities form the nodes of such a graph it can effectively model the micro-anatomy of the tissue. At the same time it is possible to make predictions at the node-, graph-, and slide-level.

Related Work. To predict the grading of colorectal cancer (CRC), both cell-based and patch-based graphs have been used in separate works [16, 23], setting the nodes of the graph as either cell nuclei or square patches, defining the node features as either handcrafted or learned features, and then applying a GNN for outcome prediction. Another patch-based GNN approach to predicting genetic mutations in CRC from H&E slides found their model trained on colon cancer generalised well to rectal cancer. For other cancers, the SlideGraph pipeline clusters nuclei for the graph nodes, and provides node-level predictions to make their model more interpretable [13]. Other approaches to setting the graph nodes include using subgraphs to represent regions [14], and creating superpatches by combining patches [11]. Edges between the nodes are usually defined by a spatial distance metric, which helps model the spatial organisation of the tissue. Common choices for GNNs include a Graph Isomorphism Network (GIN) with jumping connectivity [7, 13, 14], as we use in this research.

Our methodology proposes a novel and disease relevant approach to a more interpretable model that effectively supports a diagnostic task. Pathologists and oncologists can use this information to inspect the validity of the prediction result and interrogate key aspects of the spatial biology that is critical for patient management. Ultimately, this type of information that is not available today will help to characterise interactions between the tumour and the host tissue and therefore help to support choice of therapy. The developed framework combines self-supervised training of a Vision Transformer (ViT) to extract morphological features, a superpixel algorithm for determining nodes of a graph, and a GNN for predictions. We achieve 0.82 AUC predicting complete response to radiotherapy using deep learning on WSIs for CRC patients, whilst providing novel interpretability of the results.

2 Methods

In this section we present the patch-level feature extraction, provide the detail of the superpixel segmentation of the WSI, and illustrate the resulting graph presentation. A GNN with three branches for our output predictions is used to simultaneously make the three different predictions as shown in Fig. 1.

Pipeline. For computational reasons, all images are split into patches of size 256×256 pixels. In order to have a common feature set all the way up to the last layer of the GNN, individual patches should be represented by morphological features that are label-agnostic. This last layer of the GNN then splits into three branches to predict response to radiotherapy, the CMS4 subtype classification for CRC, and epithelial tissue regions. This way we can guarantee the common latent features and derivation across branches, maintaining the contextual importance of each branch. The DINO framework [4] uses a self-distillation training approach, using data augmentation to locally crop the patches and train with a local-global student-teacher approach. We use the DINO framework to train a ViT in a self-supervised manner on our H&E slides [6], representing each patch with 384 features. We use only the training set to train this model, and use the image patches at 20x magnification.

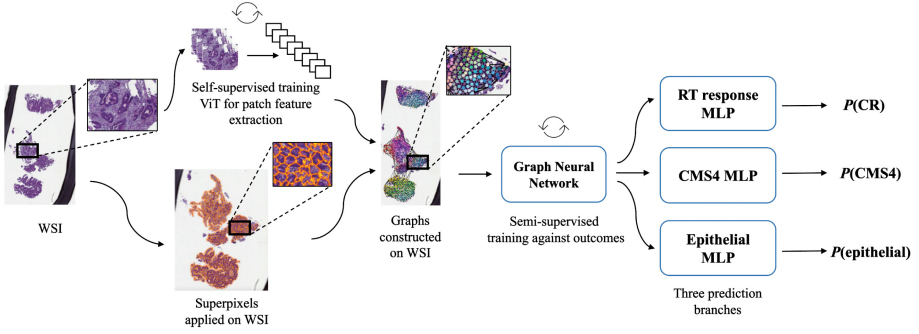


Fig. 1. Approach. We extract patch-level features from each WSI using self-supervised DINO training with a ViT model [4]. The SLIC superpixel algorithm segments the entire slide into smaller regions [1]. We calculate the mean patch features for these superpixel regions, and use the superpixel features and centers as our graph nodes, applying Delaunay triangulation to generate the edges of the graph. A GNN consisting of GINConv layers is trained on these fixed graphs, and the final layer splits into three separate MLP branches to provide predictions of three different outcomes, complete response (CR) to radiotherapy (RT), CMS4 classification, and epithelial tissue. An example output is visualized in Fig. 2.

To find the nodes of the WSI graphs, we apply the SLIC superpixel algorithm [1] on the WSIs at 5x magnification to segment the tissue to capture cellular neighbourhoods that are roughly between $80\text{--}100\text{ }\mu\text{m}^2/\text{pixels}$ in size. It can be seen that the superpixel boundaries consistently align with the boundaries of tissue compartments.

The superpixels centers are used as the nodes of the graph, and the node features are the weighted mean of the corresponding patch features which overlap with the superpixel region. The edges of the graph are determined by nearest neighbours from Delaunay triangulation, as in SlideGraph [13].

Building on the ideas introduced by SlideGraph [13] we use GINConv layers [21], adding tempering to avoid overfitting, and replace their logistic regression scaler with a simple sigmoid function. We add three branches to the final layer of the GNN, in the form of three separate multilayer perceptrons (MLPs). Two of these MLPs return a graph-level prediction, for the response to RT and CMS4 predictions, and the final branch returns node-level predictions, predicting whether each node is epithelial tissue or not. Our loss function is defined as

$$\mathcal{L} = \text{BCE}(\hat{y}_{RT}, y_{RT}) + \text{BCE}(\hat{y}_{CMS4}, y_{CMS4}) + \text{BCE}(\hat{\mathbf{y}}_{epi}, \mathbf{y}_{epi}),$$

where BCE is the binary cross entropy loss, $\hat{y}_{RT} \in \mathbb{R}$ is the slide-level prediction of response to radiotherapy, $\hat{y}_{CMS4} \in \mathbb{R}$ is the slide-level prediction of CMS4, $\hat{\mathbf{y}}_{epi} \in \mathbb{R}^{n_i}$ are the node-level predictions of epithelial tissue and n_i is the number of nodes in the i^{th} WSI graph.

For each prediction branch, we can visualize the individual node predictions from the WSI graph, overlaid on the WSI itself, to get an idea of how the node

predictions vary across the different tissue regions. Each graph-level prediction is derived from the corresponding branch node predictions, by applying pooling and dropout.

Data. We train and validate our methods on two retrospective rectal cancer datasets, Grampian and Aristotle. Both cohorts received standard chemoradiotherapy of pelvic irradiation (45–50.4Gy in 25 fractions over 5 weeks) with capecitabine 900mg/m². The pre-treatment biopsy slides were all sectioned and stained in the same laboratory, and scanned at 20x magnification (0.5 μm^2 /pixel) on an Aperio scanner. Pathological complete response, which we use as a target outcome here, was derived from histopathological assessment from post-treatment resections.

The CMS labels for this data are derived from three different transcriptomic versions (single cohort, combined cohort correcting batch effects and combined cohort including 2036 cases run with the same platform), in order to generate robust classifications. In all cases the CMS call was calculated using the CMSclassifier random forest and single sample predictor [9]. Final CMS calls are based on matching calls between the three transcriptomic versions. Despite our efforts to minimise the noise from RNA sequencing, we still expect a certain level of noise in our ground truth data, which we discuss in the Results section.

The epithelial labels for each graph node are calculated from epithelial masks for each WSI. These epithelial segmentation masks were generated at 10x magnification (1 μm^2 /pixel) with a U-Net [17] which was trained and validated on 666 full tissue sections belonging to 362 patients from the FOCUS cohort [18]. The ground truth annotations for the training of this model were generated by VK.

For consistency the tumour regions were marked up by an expert pathologist. We use these masks in our analysis to filter out background and irrelevant tissue from the images. Grampian and Aristotle are used in both training and validation, with a 70/30% training-validation split, keeping any WSIs from a single patient in the same dataset. We predict complete response to radiotherapy against all other responses, such as partial response and no response. The datasets are unbalanced, since in Grampian only 61/244 slides have complete response, and in Aristotle only 24/121 slides have complete response. They are even more unbalanced for CMS4, since only 28/244 slides in Grampian and 17/121 slides in Aristotle are labelled with CMS4. We address this imbalance in the Supplementary Materials. There are 365 slides total in our dataset, from 249 patients.

3 Experiments

Implementation. We use the default DINO parameters, but train for 20 epochs with 5 warmup epochs. We apply the SLIC algorithm [1] with compactness of 20, setting the number of segments for each WSI as half the mean size of the WSI. Prior to fitting the graph model we normalize the node features relative to the whole dataset. We train our graph model for 30 epochs using Adam

optimizer with learning rate 0.001 and weight decay 0.0001. Our graph model has three GINConv layers [21] with dimensions 64, 32 and 16 respectively. We apply dropout of 0.5 in-between graph layers, use minimum aggregation for message passing between nodes and use maximum pooling for concatenating the node activations. We apply tempering to the outcome of the graph model, dividing the output by 1.5. We evaluate the best validation epoch by finding the best mean AUC across the three prediction branches. We run the whole pipeline on four folds with different random data splits for training and validation. The code for this research will be made available upon request.

Table 1. For each fold, we take the mean metrics for the three branch predictions from the best model on our validation data, with the best epoch chosen based on mean AUC for the three predictions. The standard deviation of the metrics across the four folds is provided in brackets. Each prediction uses an optimised threshold value determined from the validation set in order to round the output probabilities to a binary prediction. We use weighted metrics due to the class imbalance in our dataset.

Response branch	Response to RT	CMS4	Epithelial
Mean AUC (std)	0.819 (0.04)	0.819 (0.04)	0.760 (0.01)
Mean accuracy (std)	0.795 (0.05)	0.750 (0.04)	0.691 (0.01)
Mean balanced accuracy (std)	0.774 (0.05)	0.719 (0.05)	0.691 (0.01)
Mean weighted F1 (std)	0.810 (0.04)	0.791 (0.02)	0.700 (0.01)
Mean weighted precision (std)	0.843 (0.02)	0.870 (0.02)	0.725 (0.00)
Mean weighted recall (std)	0.795 (0.05)	0.750 (0.04)	0.691 (0.01)

Results. Despite the noise in our reference data used for training, our model achieves good performance in terms of mean AUC scores on all three prediction branches of our model, predicting complete response to radiotherapy (RT) with 0.819 AUC, CMS4 with 0.819 AUC and epithelial tissue at the node level with 0.760 AUC across folds. Further metrics are provided in Table 1. The prediction performance of the model could be improved by utilising a larger training dataset and performing more exhaustive parameter searches, however the current performance of the model is sufficient to demonstrate the impact of this approach. The predicted response to radiotherapy can now be viewed in the context of disease biology as captured by CMS4. For example, the model demonstrates that CMS4 patients are less likely to respond to radiotherapy. In addition, it is now possible to view the spatial distribution of CMS4 active regions in the tissue architecture context as shown in Fig. 2. Additional samples are presented in the Supplementary Materials.

An example of our proposed prediction maps on two slides can be seen in Fig. 2, with further slides in the Supplementary Materials. A pathologist reviewing these maps assesses that the observed patterns fit the known interplay of response to therapy, CMS4 activation, and the spatial localisation of these signals. In the top slide, we observe high CMS4 activation in stromal rich regions,

and interestingly also high CMS4 activation in the bottom center, dissociating from the response to RT activation map. This could be explained by the lymphocyte content, supported by the higher epithelial map activations in the same location. Expert pathologists highlight a similar pattern in certain regions of the maps for the bottom slide. Different from the slide above, the CMS4 and response to RT maps have some overlap with moderate activations here, encouraging discovery into tumour-host interactions. Ultimately, a pathologist confirmed that these maps support an interpretable and trustworthy prediction in the context of response to radiotherapy. While we cannot present a more extensive interpretation of these results due to space limitations, these examples already indicate that the proposed approach enables a level of analysis that has not been possible before.

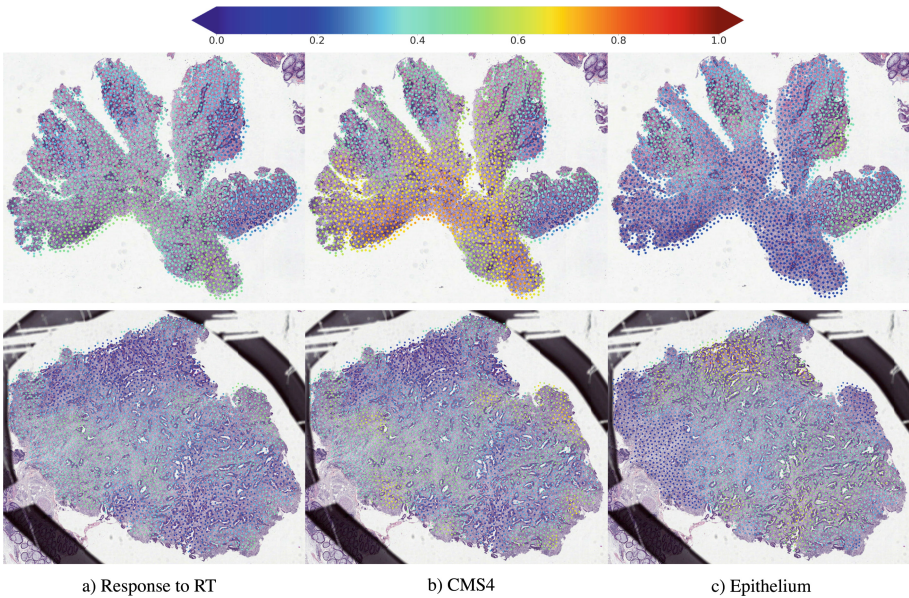


Fig. 2. Node activation maps from the three prediction branches on two different slides, top and bottom. The nodes are coloured by their predictions. Both slides are classified as CMS4 and the patients did not have a complete response to radiotherapy.

Ablation Studies. Using ablation studies, we prove our model and the prediction maps it produces are robust. Changing the dropout, loss weights, loss function, and message passing aggregation methods only changes prediction AUC scores by absolute values up to 0.03. The node activation maps are also very visually similar across ablation study models.

We find that predicting these outcomes individually in a single branch model, particularly with response to radiotherapy, can result in slightly higher AUC

scores, but we consciously make this trade-off in order to provide better interpretability of the model predictions. The focus of this research is not to achieve the best possible metrics, but to develop robust methods which can add context and explanation to clinical black box deep learning model predictions, with the view to ease clinical translation of such models.

To explore the effects of the noisy CMS4 ground truth labels, we remove from our dataset any WSIs classified as ‘Unmatched’ for the CMS call, which for the main results of this paper we defined as ‘Not CMS4’. Removing this data and rerunning our analysis improved our predictions for CMS4 by +0.06 AUC, and reduced our response to radiotherapy and epithelial predictions by −0.02 and −0.01 respectively. The results can be found in the Supplementary Materials. These small changes indicate that the noise in our data does not degrade the performance of our classifier, reinforcing it as a robust and accurate model.

4 Conclusion

By setting the prediction of response to therapy in context with disease biology and spatial organisation of the tissue we are providing a novel approach for enhancing the interpretability of complex prediction tasks. These results do not only enhance the interpretability, they also provide new ways to utilise large retrospective clinical trial cohorts for which no additional molecular data is available. Extending the amount of training data and improving model training will improve model performance, which is already impressive.

We argue that this work also advances the state of the art in feature representation and analysis. Our prediction maps derive from the same graph model, and hence they share underlying graph features. The prediction branches only diverge at the final stage of translating these graph features into outcome predictions for our three clinically relevant outcomes. Importantly, this level of visualisation is not only accessible to pathologists, this joint prediction model also enhances the communication between pathologists and oncologists which is critical for patient management. By cross-referencing these prediction maps with our prior understanding of cancer biology, this approach can help to establish trust in the prediction model and also help to identify potential failure cases.

This work relies on access to well annotated clinical trial samples which will limit our ability to include more data for training and testing. In future, we plan to use these methods to help better characterise tumour-stromal interactions of the tissue. We also plan to use a denser graph with less connectivity to be able to better predict the heterogeneous epithelial tissue.

Data Use Declaration and Acknowledgements. The Grampian, Aristotle and FOCUS datasets used in this study were available through the S:CORT consortium. Grampian and Aristotle will be made publicly available shortly. Additional data is available on request to the S:CORT consortium. The S:CORT consortium was reviewed and approved by the South Cambs Research Ethics committee (REC ref 15/EE/0241; IRAS reference 169363). The Stratification in Colorectal Cancer Consortium (S:CORT) was funded by the Medical Research Council and Cancer Research UK (MR/M016587/1).

The Aristotle trial was funded by Cancer Research UK (CRUK/08/032). The funders played no role in the analyses performed or the results presented. Financial support: RW - EPSRC Center for Doctoral Training in Health Data Science (EP/S02428X/1), Oxford CRUK Cancer Centre; VHK - Promedica Foundation (F-87701-41-01) and Swiss National Science Foundation (P2SKP3_168322/1, P2SKP3_168322/2); TSM - S:CORT (see above); JR, KS - Oxford NIHR National Oxford Biomedical Research Centre and the PathLAKE consortium (InnovateUK). The computational aspects of this research were funded from the NIHR Oxford BRC with additional support from the Wellcome Trust Core Award Grant Number 203141/Z/16/Z. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(11), 2274–2282 (2012)
2. Alkan, A., Hofving, T., Angenete, E., Yrlid, U.: Biomarkers and cell-based models to predict the outcome of neoadjuvant therapy for rectal cancer patients. *Biomark. Res.* **9**(1) (2021)
3. Anitei, M.G., et al.: Prognostic and predictive values of the immunoscore in patients with rectal cancer. *Clin. Cancer Res.* **20**(7), 1891–1899 (2014)
4. Caron, M., et al.: Emerging properties in self-supervised vision transformers. In: *Proceedings of the International Conference on Computer Vision (ICCV)* (2021)
5. Chatila, W., Kim, J., Walch, H.: Genomic and transcriptomic determinants of response to neoadjuvant therapy in rectal cancer. *Nat. Med.* **28**, 1646–1655 (2022)
6. Chen, R.J., Krishnan, R.G.: Self-supervised vision transformers learn visual concepts in histopathology (2022)
7. Ding, K., Liu, Q., Lee, E., Zhou, M., Lu, A., Zhang, S.: Feature-enhanced graph networks for genetic mutational prediction using histopathological images in colon cancer. In: Martel, A.L., et al. (eds.) *MICCAI 2020. LNCS*, vol. 12262, pp. 294–304. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59713-9_29
8. George, T.J.J., Allegra, C.J., Yothers, G.: Neoadjuvant rectal (NAR) score: a new surrogate endpoint in rectal cancer clinical trials. *Curr. Colorect. Cancer Rep.* **11**(5), 275–280 (2015)
9. Guinney, J., et al.: The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015)
10. Jones, H.J.S., et al.: Stromal composition predicts recurrence of early rectal cancer after local excision. *Histopathology* **79**, 947–956 (2021)
11. Lee, Y., Park, J., Oh, S.: Derivation of prognostic contextual histopathological features from whole-slide images of tumours via graph deep learning. *Nat. Biomed. Eng.* (2022)
12. Lipkova, J., et al.: Artificial intelligence for multimodal data integration in oncology. *Cancer Cell* **40**(10), 1095–1110 (2022)
13. Lu, W., Toss, M., Dawood, M., Rakha, E., Rajpoot, N., Minhas, F.: Slidegraph+: whole slide image level graphs to predict her2 status in breast cancer. *Med. Image Anal.* **80**, 102486 (2022)

14. Pina, O., Vilaplana, V.: Self-supervised graph representations of WSIS. In: Bekkers, E., Wolterink, J.M., Aviles-Rivero, A. (eds.) *Proceedings of the First International Workshop on Geometric Deep Learning in Medical Image Analysis. Proceedings of Machine Learning Research*, vol. 194, pp. 107–117. PMLR (2022)
15. Qi, L., et al.: Identification of prognostic spatial organization features in colorectal cancer microenvironment using deep learning on histopathology images. *Med. Omics* **2**, 100008 (2021)
16. Raju, A., Yao, J., Haq, M.M.H., Jonnagaddala, J., Huang, J.: Graph attention multi-instance learning for accurate colorectal cancer staging. In: Martel, A.L., et al. (eds.) *MICCAI 2020. LNCS*, vol. 12265, pp. 529–539. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59722-1_51
17. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation (2015)
18. Sirinukunwattana, K., et al.: Image-based consensus molecular subtype (imcms) classification of colorectal cancer using deep learning. *Gut* **70**(3), 544–554 (2021)
19. Wood, R., et al.: Enhancing local context of histology features in vision transformers. In: *Artificial Intelligence over Infrared Images for Medical Applications and Medical Image Assisted Biomarker Discovery*. pp. 154–163. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-19660-7_15
20. Xu, J., Luo, X., Wang, G., Gilmore, H., Madabhushi, A.: A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images. *Neurocomputing* **191**, 214–223 (2016)
21. Xu, K., Hu, W., Leskovec, J., Jegelka, S.: How powerful are graph neural networks? (2018)
22. Zhang, F., et al.: Predicting treatment response to neoadjuvant chemoradiotherapy in local advanced rectal cancer by biopsy digital pathology image features. *Clin. Transl. Med.* **10**(2), e110 (2020)
23. Zhou, Y., Graham, S., Koohbanani, N.A., Shaban, M., Heng, P.A., Rajpoot, N.M.: CGC-net: cell graph convolutional network for grading of colorectal cancer histology images. In: *IEEE/CVF International Conference on Computer Vision Workshops*, pp. 388–398. IEEE (2019)