Enhancing Local Context of Histology Features in Vision Transformers *

Ruby Wood[1,3][0000−0001−9916−888X], Korsuk Sirinukunwattana[1,3][0000−0002−3603−4435], Enric Domingo[4][0000−0003−4390−8767], Alexander Sauer[1,3][0000−0003−1563−6143], Maxime W Lafarge[5], Viktor H Koelzer[5][0000−0001−9206−4885], Timothy S Maughan[4][0000−0002−0580−5065], and Jens Rittscher[1,2,3](✉)[0000−0002−8528−8298]

[1] Department of Engineering Science, University of Oxford, UK
{ruby.wood|jens.rittscher}@eng.ox.ac.uk
[2] Nuffield Department of Medicine, University of 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 CH-8091, Switzerland

Abstract. Predicting complete response to radiotherapy in rectal cancer patients using deep learning approaches from morphological features extracted from histology biopsies provides a quick, low-cost and effective way to assist clinical decision making. We propose adjustments to the Vision Transformer (ViT) network to improve the utilisation of contextual information present in whole slide images (WSIs). Firstly, our position restoration embedding (PRE) preserves the spatial relationship between tissue patches, using their original positions on a WSI. Secondly, a clustering analysis of extracted tissue features explores morphological motifs which capture fundamental biological processes found in the tumour micro-environment. This is introduced into the ViT network in the form of a cluster label token, helping the model to differentiate between tissue types. The proposed methods are demonstrated on two large independent rectal cancer datasets of patients selectively treated with radiotherapy and capecitabine in two UK clinical trials. Experiments demonstrate that both models, PREViT and ClusterViT, show improvements in the prediction over baseline models.

Keywords: Vision Transformer · Histology · Whole Slide Image · Clustering.

1 Introduction

Around 11,500 people are diagnosed each year with rectal cancer in the UK [22]. Neoadjuvant treatment in the form of radiotherapy is commonly given to patients

* Supported by Cancer Research UK.

with locally advanced disease to shrink tumour prior to surgery. However, one third of rectal cancer patients do not benefit from this treatment [9]. Determining patient response to radiotherapy therefore is critical to avoid overtreatment.

The following studies provide a first indication that biological features of the primary tumour can be captured in biopsy fragments and allow prediction of clinical disease behaviour. Koelzer et al. [18] show that CD8 infiltration in rectal cancer biopsies before surgery associates with favourable outcome, and Anitei et al. [2] propose that tumour immune infiltrate evaluated with their Immunoscore method is a useful prognostic marker for rectal cancer patients prior to surgery. Furthermore, Jones et al. [15] find that increased stromal content in early rectal cancer is a predictor for an increased risk of recurrence, and Rogers et al. [23] find that tumour budding in the biopsy taken prior to chemoradiotherapy is a negative predictor of pathological response. These studies, however, focus on single features rather than visual assessment, and are based on limited cohorts with heterogeneous treatment conditions.

Recent research has demonstrated the application of deep learning to predict a patient's response to therapy and to predict biologically relevant information from morphological features extracted from standard histology slides. For example, microsatellite instability (MSI) is one of the key markers informing the treatment decision in colorectal cancer (CRC) [17, 1]. Research has shown it is possible to predict MSI status using deep learning from standard hematoxylin and eosin (H&E) stained slide images [3, 7, 11]. As well as MSI, Bilal et al. [3] predict some other key molecular pathways and mutations such as chromosomal instability, CpG island methylator phenotype and tumour-infiltrating lymphocytes from CRC histology slides. Sirinukunwattana et al. [26] have proven that the consensus molecular subtypes (CMS) of CRC [10], a stratification system with clear biological interpretation, can be predicted from H&E images using deep learning approaches. Although Zhang et al. [27] have demonstrated the ability to predict the chemoradiotherapy response in locally advanced rectal cancer from H&E biopsies, they only used standard machine learning approaches.

Applying deep learning to whole slide images (WSIs) of histology samples is not computationally straightforward owing to their large pixel numbers. Weakly supervised techniques are commonly used, whereby each WSI is split into patches or tiles. Each patch then receives its own label from the parent slide and subsequently receives its own output which can be aggregated into the output of the parent slide. Various patch prediction aggregation methods can be used. Campanella et al. [5] and others [4, 12, 16] explore max-pooling methods and recurrent neural networks for the aggregation.

To date, most deep learning approaches utilise this tile-based approach, aggregating independent tile predictions to obtain a slide-level prediction. We contend that there is a benefit to integrating local context information as early as possible in the learning process. Using the attention mechanisms of the Vision Transformer (ViT) model [6] to get a slide-level prediction provides an effective way to obtain this goal. Li et al. [19] propose to use a ViT to predict the slide-level outcome from patch features extracted from an EfficientNet B0 (pre-trained on ImageNet). Lu et al. [21] develop the CLAM model (Clustering-Constrained-Attention Multiple-instance learning) which combines clustering analysis and attention mechanism to aggregate their patch predictions. They first train a convolutional neural network (CNN) to extract features for the image patches and then apply an attention-based pooling function [13] to generate slide-level predictions. Clustering of patch features is used to constrain the feature space and refine predictions. Sharma et al. [25] use the K-Means clustering technique to cluster and sample patches extracted from a CNN, on which they apply the attention-based pooling function [13] for a slide-level prediction.

In this work, we will demonstrate the ability of the state-of-the-art ViT models to predict response to radiotherapy in locally advanced rectal cancer patients, which to our knowledge have never been applied for this purpose before, and will demonstrate their ability to provide visual interpretation to image features associated with the treatment outcomes. We propose two novel variations of ViT to enhance the local context of the tissue patch features by introducing prior information to the model. The first network is the PREViT which uses a novel trainable position restoration embedding (PRE) to restore patches to their original position in the WSI, preserving the spatial relationship between tissue patches. The second is the ClusterViT which introduces a trainable cluster label encoding derived from the baseline model feature embeddings as prior information for the ViT model to capture the different tissue motifs. Two independent clinical trial cohorts (see datasets in Section 3 for a more detailed description) are used in the development and validation of the models. Experiments demonstrate that the proposed models improve predictions over the baseline model and the interpretation of predictions is aided by including the PRE and cluster token.

2 Methods

First, an outline of the baseline model is given. Subsequently, we introduce how the ViT model is extended with PRE. We also introduce the ClusterVIT, where we specify how relevant morphological motifs are found through standard clustering methods. Finally, we combine the two proposed methods into the ClusterPREViT model. An overview of the resulting imaging pipeline is shown in Fig. 1.

Baseline model We first split the WSIs into smaller patches of size 256 x 256 pixels and use tumour masks provided by a pathologist to filter out unwanted tissue and background. We train a ResNet-50 model (pre-trained on ImageNet) to predict patch labels, using the slide-level outcome as a label for each patch in that slide. Once training is complete we use a mean-pooling approach, taking the mean over the predictions of all patches within a slide for a slide-level prediction. In addition, we extract feature embeddings $\in \mathbb{R}^{2048}$ from the penultimate layer of the ResNet model for use in further models.

PREViT Shao et al. [24] propose the PPEG (Pyramid Position Encoding Generator) module, designed to capture positional information of the tissue

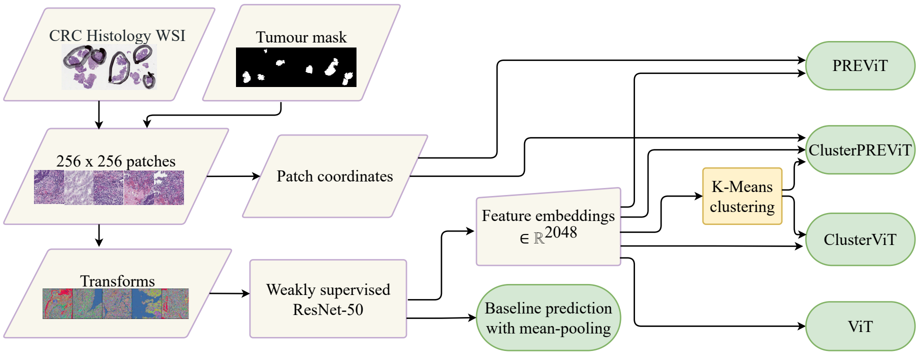

Fig. 1. Overview of the full model pipeline, showing the baseline model and the ViT model variants. The tumour masks are applied to the WSIs to filter out the background and non-tumour tissue. The remaining tissue regions are then split into patches to feed into a ResNet-50, which generates patch feature embeddings and a baseline prediction from mean-pooling patch predictions across slides. The features are clustered with K-Means to provide the cluster token as contextual prior information in the ClusterViT model, and the patch coordinates are used to preserve the spatial relationship of the tissue patches in the PREViT model. Both cluster labels and patch coordinates are given to the combined ClusterPREViT model.

patches. They first restore the input sequence into a 2D image space and then apply CNNs to encode the position information, using convolution kernels of different sizes to capture spatial information with different granularity. However, they convert the input sequence to lie in a 2D image space without considering the original position of the patches in the WSI. This method does not apply well when the patches originate from different parts of the WSI such as different parts of tissue, and when background patches have been excluded. Therefore, naively converting the input into 2D space is not contextually meaningful.

We propose a novel patch PRE, which restores the original position of the patch relative to the WSI, using zero padding for areas of the image which do not qualify for training [14], and applies convolutions in the same manner as the PPEG module. The motivation for the patch PRE is to incorporate local information around the patch, acting as a prior in the model representing the surrounding tissue environment.

ClusterViT To capture the different tissue motifs across the WSIs we perform a clustering analysis, using a K-Means clustering model on the feature embeddings from the baseline model. The purpose of the clustering is to detect the morphological patterns that capture interactions between tissue types including stroma, epithelium and immune cells. The training set (used for training the baseline ResNet model) is used to estimate the parameters of the K-Means model which is then used to predict the cluster labels for the validation set of slides. The K-Means model is learnt on the L_2-normalised feature embeddings, trained across the whole training set (as opposed to per slide). Each patch receives its own cluster label. The number of clusters is chosen to be $k = 4$, inspired by the marked morphological correlation to the four molecular subtypes found in Sirinukunwattana et al. [26]. Alternative values for the numbers of clusters, k, were considered in the development process.

We propose a novel ViT network, ClusterViT, which incorporates a learnable cluster token which is added to the input after the classification token is appended but before it is passed to the main Transformer attention network in a similar vein to the position embedding. This idea is inspired by Gao et al. [8], who add a nuclei grade embedding to their ViT model to capture prior information on the nuclei for subtyping of papillary renal cell carcinoma. We encode the cluster labels using an encoding from Gao et al. [8]. The aim of the cluster token is to consider the tissue morphology for each patch in the WSI, and use this as prior information for the model prediction. The cluster label from each patch, derived from the aforementioned K-Means clustering, is embedded and concatenated to the input as a learnable cluster token.

ClusterPREViT The ClusterPREViT model is a ViT model combining both the patch PRE and the cluster token from the ClusterViT model. The ViT classification token is appended to the input first, followed by the cluster token and finally the PRE.

3 Experiments

Datasets The proposed methods are applied on H&E stained histology slides from two retrospective rectal cancer datasets, Grampian and Aristotle (each from separate UK clinical trials). All patients were selected to receive standard chemoradiotherapy comprising pelvic irradiation (45-50.4Gy in 25 fractions over 5 weeks) with capecitabine 900mg/m^2. Pathological complete response was assessed by detailed histopathological assessment of the resection specimens after treatment. All slides from pre-treatment biopsies had been sectioned and stained in the same laboratory and scanned at high resolution on an Aperio scanner at a total magnification of 20x (0.5 μm^2/pixel), according to criteria from the Royal College of Pathologists.

In the Grampian dataset, 258 WSIs from 133 patients have recorded response to radiotherapy after data quality control steps. In the Aristotle dataset, 124 WSIs from 124 patients have response to radiotherapy after quality control. The patient's response to treatment is split into two classes across the trial cohorts, measured as either a complete response or no complete response to radiotherapy. Both trial datasets are unbalanced, with a total of 91 slides labelled as complete response and 291 labelled as no complete response to radiotherapy. All WSIs in this analysis have a tumour mask generated by hand by a pathologist. For all modelling we split both datasets for training and validation such that all slides from a single patient are in the same dataset, with a ratio of 70% training vs 30% validation.

In the ResNet training process, the final layer of the ResNet model is set to return a single value, using a linear layer with 2048 features as input and one

Table 1. Validation results from running each model for five rounds using different random seeds and different data splits across the two combined datasets determined by the random seed. The mean weighted area under the curve of the receiver operating characteristic curve (AUC) and standard deviation (std) over the five rounds are presented as the primary metric, as well as accuracy, F1 score, precision and recall, all weighted by class-balanced sample weights.

Model	Mean AUC (std)	Accuracy	F1 Score	Precision	Recall
ResNet	0.696 (0.116)	0.696	0.787	0.802	0.788
ViT	0.857 (0.017)	0.725	0.823	0.829	0.829
PREViT	0.859 (0.021)	0.754	**0.834**	0.844	**0.836**
ClusterViT	0.859 (0.028)	**0.767**	0.781	**0.845**	0.767
ClusterPREViT	**0.861** (0.013)	0.763	0.782	0.843	0.768

feature as output. For all our experiments, we use the patched images at 10x magnification (1 μm^2/pixel). We use stochastic gradient descent as the optimiser function with a learning rate of 0.1, momentum of 0.9 and weight decay of 1e-4. The models train on a batch size of 256 patches. The ResNet models are trained for 20 epochs. For all ViT models, we train and validate on the same data as for the corresponding baseline model. We use stochastic gradient descent as the optimiser function with a learning rate of 0.001, momentum of 0.9 and weight decay of 1e-4. The models train on all patch features for one slide at once. In the case when the number of patches per slide is unusually large, for computational reasons we randomly select 10,000 patches from the slide to train or validate on. All ViT models are trained for 50 epochs.

Results We run the full model pipeline for five rounds, using different random seeds and different data splits across the datasets, determined by the random seed. The results presented in Table 1 are the metrics from the five rounds, all weighted by class-balanced sample weights due to the dataset imbalance. The default threshold of 0.5 is used for the metrics which require a binarised prediction. This provides a valid technical comparison, but other metrics and thresholds for clinical translation will be explored in future work. The AUC is used as the primary metric here since it is more discriminating and consistent than the accuracy score [20].

In terms of our primary metric the best performing model is the ClusterPREViT, with a weighted AUC of 0.861 over five rounds and the lowest standard deviation in AUC. The PREViT and ClusterViT models perform marginally better than the ViT, but all show a substantial improvement in performance over the baseline ResNet model, across all measured metrics.

The benefit of both the clustering and position embedding approaches are that they provide more insight into the model and the model predictions. The clustering assigns a cluster label to each patch which can be visualised on the WSI as prior information for the attention model, and the PREViT model provides particularly informative attention heatmaps. Both approaches provide clues to morphological continuity in a slide which suppresses the ViT models from attending to small spurious regions and encourages the attention towards regions with

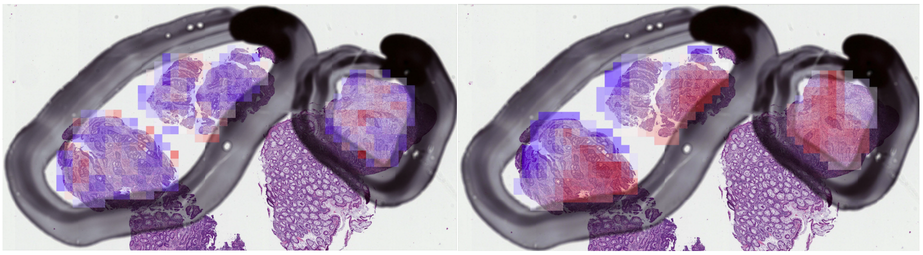

Fig. 2. Heatmaps showing the attention weights from the model predictions for this slide. On the left is the heatmap from the ViT and on the right is the PREViT heatmap. A pathologist reviewing these heatmaps observed that the PREViT model demonstrates more attention to areas of invasive cancer, and less attention to non-informative artefacts.

morphological meaning. As such, both approaches increase the interpretability of the heatmaps of the attention weights from the ViT model predictions. Fig. 2 shows the comparison of a heatmap of the PREViT model attention weights for one slide (right) against the ViT model attention weights (left). These heatmaps can provide feedback to pathologists for a visual interpretation of the model results, though further exploration and model optimisation is required to determine whether the attention heatmaps could be informative in a clinical setting.

In order to demonstrate the clinical relevance of our prediction, we carry out a multivariate logistic regression for the validation set of each round in order to determine the odds ratio that a patient responds to radiotherapy. In this model, we find that the prediction made by the ClusterPREViT is the most predictive covariate, since the confidence intervals for the estimated effect are the furthest away from zero. This holds in all five rounds when compared to T stage, N stage, age and gender as shown in Fig. 3.

4 Conclusion

The proposed extensions to the ViT framework demonstrate the potential of predicting response to radiotherapy from features extracted from standard H&E slides. As such we are presenting a new and exciting application of computational pathology. Expanding the ViT model effectively enhances performance of the prediction as well as capturing spatial information, by capitalising on morphological tissue clusters and spatial positioning on the WSI. Clustering the tissue features enhances the ability to provide visual feedback which ultimately makes our approach more usable in clinical translation. Future work will address optimising a threshold for the binary response classification for use in the clinical setting, and will further utilise the morphology of different tissue components to improve model predictions and interpretations.

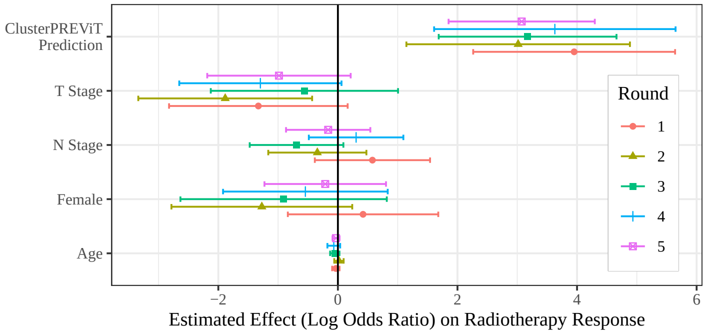

Fig. 3. Log odds ratios for covariates in a logistic regression model predicting complete response to radiotherapy. The most predictive covariate is the ClusterPREViT model prediction when compared to T stage, N stage, age and gender. The validation set for each round was used in a separate model, hence five lines for each covariate.

Acknowledgements Financial support: RW, AS - EPSRC Center for Doctoral Training in Health Data Science (EP/S02428X/1); RW - Oxford CRUK Cancer Centre; VHK - Promedica Foundation (F-87701-41-01) and Swiss National Science Foundation (P2SKP3_168322/1, P2SKP3_168322/2); TSM - The Stratification in Colorectal Cancer Consortium (S:CORT) funded by the Medical Research Council and Cancer Research UK (MR/M016587/1); JR, KS - Oxford NIHR National Oxford Biomedical Research Centre and the PathLAKE consortium (InnovateUK). The ARISTOTLE trial was funded by Cancer Research UK (CRUK/08/032). 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. André, T., Shiu, K., Kim, T., et al.: Pembrolizumab in microsatellite-instability-high advanced colorectal cancer. The New England Journal of Medicine **383**(23), 2207–2218 (2020), doi:10.1056/NEJMoa2017699
2. Anitei, M.G., Zeitoun, G., Mlecnik, B., Marliot, F., Haicheur, N., Todosi, A.M., Kirilovsky, A., Lagorce, C., Bindea, G., Ferariu, D., Danciu, M., Bruneval, P., Scripcariu, V., Chevallier, J.M., Zinzindohoué, F., Berger, A., Galon, J., Pagès, F.: Prognostic and predictive values of the immunoscore in patients with rectal cancer. Clinical Cancer Research **20**(7), 1891–1899 (April 2014), https://doi.org/10.1158/1078-0432.CCR-13-2830
3. Bilal, M., Raza, S., Azam, A., et al.: Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: a retrospective study. Lancet Digit Health **3**(12), e763–e772 (2021)

4. Bychkov, D., Linder, N., Turkki, R., et al.: Deep learning based tissue analysis predicts outcome in colorectal cancer. Scientific Reports **8**, 3395 (2018), https://doi.org/10.1038/s41598-018-21758-3
5. Campanella, G., Hanna, M., Geneslaw, L., et al.: Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature Medicine **25**, 1301–1309 (2019), https://doi.org/10.1038/s41591-019-0508-1
6. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. CoRR (2020), https://arxiv.org/abs/2010.11929
7. Echle, A., Grabsch, H.I., Quirke, P., et al.: Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning. Gastroenterology **159**(4), 1406–1416 (2020)
8. Gao, Z., Hong, B., Zhang, X., Li, Y., Jia, C., Wu, J., Wang, C., Meng, D., Li, C.: Instance-based vision transformer for subtyping of papillary renal cell carcinoma in histopathological image. CoRR (2021), https://arxiv.org/abs/2106.12265
9. George, T.J.J., Allegra, C.J., Yothers, G.: Neoadjuvant rectal (nar) score: a new surrogate endpoint in rectal cancer clinical trials. Current Colorectal Cancer Reports **11**(5), 275–280 (2015), doi:10.1007/s11888-015-0285-2
10. Guinney, J., Dienstmann, R., Wang, X., et al.: The consensus molecular subtypes of colorectal cancer. Nature Medicine **21**, 1350–1356 (2015), https://doi.org/10.1038/nm.3967
11. Hildebrand, L.A., Pierce, C.J., Dennis, M., Paracha, M., Maoz, A.: Artificial intelligence for histology-based detection of microsatellite instability and prediction of response to immunotherapy in colorectal cancer. Cancers (Basel) **13**(3), 391 (2021), doi:10.3390/cancers13030391
12. Iizuka, O., Kanavati, F., Kato, K., et al.: Deep learning models for histopathological classification of gastric and colonic epithelial tumours. Scientific Reports **10**, 1504 (2020), https://doi.org/10.1038/s41598-020-58467-9
13. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: Dy, J., Krause, A. (eds.) Proceedings of the 35th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 80, pp. 2127–2136. PMLR (2018), https://proceedings.mlr.press/v80/ilse18a.html
14. Islam, M.A., Jia, S., Bruce, N.D.B.: How much position information do convolutional neural networks encode? CoRR (2020), https://arxiv.org/abs/2001.08248
15. Jones, H.J.S., Cunningham, C., Askautrud, H.A., Danielsen, H.E., Kerr, D.J., Domingo, E., Maughan, T., Leedham, S.J., Koelzer, V.H.: Stromal composition predicts recurrence of early rectal cancer after local excision. Histopathology **79**, 947–956 (2021), https://doi.org/10.1111/his.14438
16. Kanavati, F., Toyokawa, G., Momosaki, S., et al.: A deep learning model for the classification of indeterminate lung carcinoma in biopsy whole slide images. Scientific Reports **11**, 8110 (2021), https://doi.org/10.1038/s41598-021-87644-7
17. Kim, N., Kim, S.M., Lee, B.J., Choi, B.i., Yoon, H.S., Kang, S.H., Kim, S.H., Joo, M.K., Park, J.J., Kim, C.: Detection of microsatellite instability in colorectal cancer patients with a plasma-based real-time pcr analysis. Frontiers in Pharmacology **12** (2021). https://doi.org/10.3389/fphar.2021.758830, https://www.frontiersin.org/article/10.3389/fphar.2021.758830
18. Koelzer, V., Lugli, A., Dawson, H., et al.: Cd8/cd45ro t-cell infiltration in endoscopic biopsies of colorectal cancer predicts nodal metastasis and survival. Journal of Translational Medicine **12**(81) (2014), https://doi.org/10.1186/1479-5876-12-81

19. Li, H., Yang, F., Zhao, Y., Xing, X., Zhang, J., Gao, M., Huang, J., Wang, L., Yao, J.: Dt-mil: Deformable transformer for multi-instance learning on histopathological image. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. pp. 206–216. Springer International Publishing (2021)
20. Ling, C.X., Huang, J., Zhang, H.: Auc: A better measure than accuracy in comparing learning algorithms. In: Xiang, Y., Chaib-draa, B. (eds.) Advances in Artificial Intelligence. pp. 329–341. Springer Berlin Heidelberg, Berlin, Heidelberg (2003)
21. Lu, M.Y., Williamson, D., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature Biomedical Engineering **5**(6), 555–570 (2021), doi:10.1038/s41551-020-00682-w
22. New treatment could spare early-stage rectal cancer patients life-altering side effects. https://www.birmingham.ac.uk/news/2020/new-treatment-could-spare-early-stage-rectal-cancer-patients-life-altering-side-effects, accessed: 21-06-2022
23. Rogers, A., Gibbons, D., Hanly, A., et al.: Prognostic significance of tumor budding in rectal cancer biopsies before neoadjuvant therapy. Modern Pathology **27**, 156–162 (2014), https://doi.org/10.1038/modpathol.2013.124
24. Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., Zhang, Y.: Transmil: Transformer based correlated multiple. instance learning for whole slide image classication. CoRR (2021), https://arxiv.org/abs/2106.00908
25. Sharma, Y., Shrivastava, A., Ehsan, L., Moskaluk, C.A., Syed, S., Brown, D.E.: Cluster-to-conquer: A framework for end-to-end multi-instance learning for whole slide image classification (2021), doi.org/10.48550/arxiv.2103.10626
26. Sirinukunwattana, K., Domingo, E., Richman, S.D., Redmond, K.L., Blake, A., Verrill, C., Leedham, S.J., Chatzipli, A., Hardy, C., Whalley, C.M., Wu, C.h., Beggs, A.D., McDermott, U., Dunne, P.D., Meade, A., Walker, S.M., Murray, G.I., Samuel, L., Seymour, M., Tomlinson, I., Quirke, P., Maughan, T., Rittscher, J., Koelzer, V.H.: Image-based consensus molecular subtype (imcms) classification of colorectal cancer using deep learning. Gut **70**(3), 544–554 (2021). https://doi.org/10.1136/gutjnl-2019-319866, https://gut.bmj.com/content/70/3/544
27. Zhang, F., Yao, S., Li, Z., et al.: Predicting treatment response to neoadjuvant chemoradiotherapy in local advanced rectal cancer by biopsy digital pathology image features. Clinical and Translational Medicine **10**(2), e110 (2020), doi:10.1002/ctm2.110