



Whole Slide Images are 2D Point Clouds: Context-Aware Survival Prediction Using Patch-Based Graph Convolutional Networks

Richard J. Chen^{1,2,3,4}(✉), Ming Y. Lu^{1,2,3,4}, Muhammad Shaban^{1,2,3,4},
Chengkuan Chen^{1,2,3,4}, Tiffany Y. Chen^{1,2,3,4}, Drew F. K. Williamson^{1,2,3,4},
and Faisal Mahmood^{1,2,3,4}

¹ Department of Pathology, Brigham and Women's Hospital, Boston, USA
faisalmahmood@bwh.harvard.edu

² Department of Biomedical Informatics, Harvard Medical School, Boston, USA
richardchen@g.harvard.edu

³ Cancer Data Science Program, Dana-Farber Cancer Institute, Boston, USA

⁴ Cancer Program, Broad Institute of Harvard and MIT, Cambridge, USA

Abstract. Cancer prognostication is a challenging task in computational pathology that requires context-aware representations of histology features to adequately infer patient survival. Despite the advancements made in weakly-supervised deep learning, many approaches are not context-aware and are unable to model important morphological feature interactions between cell identities and tissue types that are prognostic for patient survival. In this work, we present Patch-GCN, a context-aware, spatially-resolved patch-based graph convolutional network that hierarchically aggregates instance-level histology features to model local- and global-level topological structures in the tumor microenvironment. We validate Patch-GCN with 4,370 gigapixel WSIs across five different cancer types from the Cancer Genome Atlas (TCGA), and demonstrate that Patch-GCN outperforms all prior weakly-supervised approaches by 3.58–9.46%. Our code and corresponding models are publicly available at <https://github.com/mahmoodlab/Patch-GCN>.

Keywords: Computer vision · Computational pathology ·
Weakly-supervised learning · Graph convolutional networks ·
Interpretability

1 Introduction

Weakly-supervised deep learning has made remarkable progress in computational pathology in using whole slide images (WSIs) for cancer diagnosis and prognosis [1–5]. Due to the computational complexities in training with WSIs, many

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-87237-3_33) contains supplementary material, which is available to authorized users.

weakly-supervised methods have approached WSIs using multiple instance learning (MIL), in which: 1) small image patches from the WSI are extracted as independent instances, and then 2) pooled using a global aggregation operator over the bag of unordered instances. Despite not being context-aware and without needing detailed clinical annotation, many of these MIL-based approaches are able to still solve difficult tasks such as cancer grading and subtyping using only slide-level labels, as the distinction between morphological phenotypes such as tumor vs. non-tumor tissue may only depend on instance-level patch-based features [6].

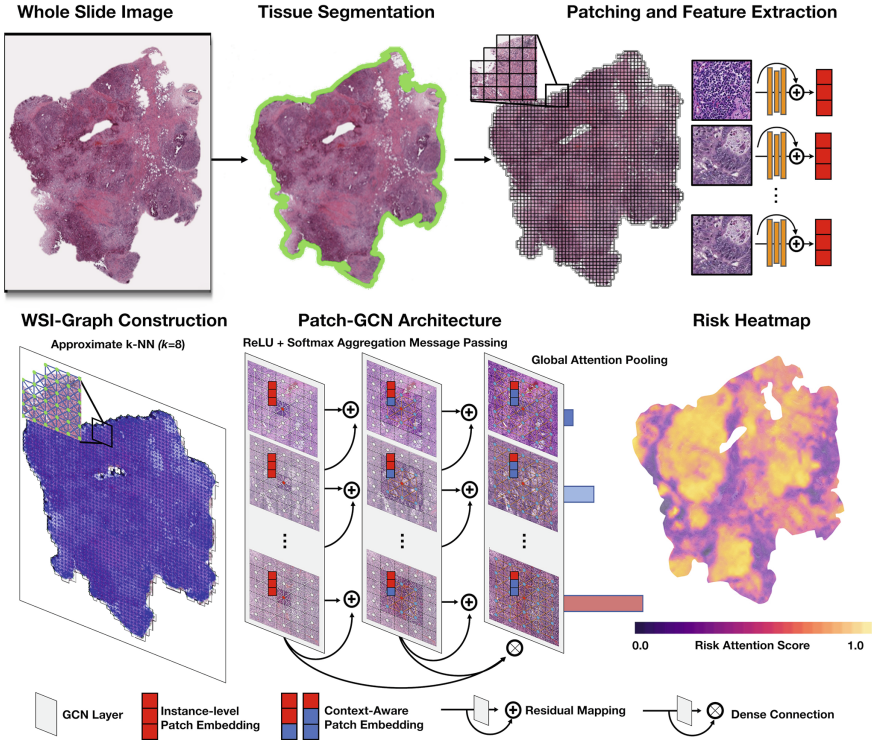


Fig. 1. Patch-GCN framework for context-aware survival outcome prediction in WSIs. Non-overlapping 256×256 patches are patched as used as input into a ResNet-50 CNN to construct the node feature matrix, with edges drawn between adjacent image patches in the WSI. A ReLU + Softmax Message Passing scheme is used to aggregate instance-level embeddings in local neighborhoods, with residual mappings and skip connections used to construct context-aware embeddings, followed by global attention-based pooling.

In contrast with cancer grading and subtyping, cancer prognostication is a challenging task that requires considering both instance- and global-level features in the tumor and surrounding tissues for assessing patient risk of mortality [7, 8]. In adapting the MIL framework to WSIs, many approaches follow the standard multiple instance (SMI) assumption for solving clinical tasks in computational pathology, *e.g.* - if a bag contains at least one positive instance, it is labeled positive,

else negative [9]. This assumption holds when the clinical task is solving binary instance-level feature discrimination problems such as tumor vs. non-tumor tissue. However, in tasks such as survival outcome prediction in cancer pathology, MIL-based approaches are unable to capture important contextual and hierarchical information that have known prognostic significance in cancer survival [10, 11]. For example, though MIL would be able to learn instance-level features that discriminate image patches of lymphocytes and tumor cells, it is unable to distinguish whether those immune cells are tumor-infiltrating lymphocytes (TILs) or from an adjacent inflammatory response, which depends on the lymphocytes' apposition to tumor cells or normal stroma respectively [8, 12].

In this work, we propose a context-aware, spatially-resolved patch-based graph convolutional network (Patch-GCN) for survival prediction in patients with multiple WSIs (Fig. 1). One of the key contributions of our work is that we formulate WSIs as a graph-based data structure in the Euclidean space similar to a point cloud in which: 1) nodes correspond to histology image patches, and 2) edges are connected between adjacent image patches from the true spatial coordinates of the WSI. As a result, message passing in Patch-GCN generalizes the standard convolutional operator in CNNs, in which node features are hierarchically aggregated from local to global structures in the WSI. Compared to other weakly-supervised learning approaches such as MIL, Patch-GCN is context-aware and is able to build hierarchical representations of morphological image patch features in context with their surrounding environment. To robustly validate Patch-GCN, we quantitatively assessed our model on five different cancer datasets from The Cancer Genome Atlas (TCGA) in survival outcome prediction against several state-of-the-art methods in weakly-supervised learning for WSIs, and evaluated the interpretability of Patch-GCN through attention heatmaps in low and high risk patients (Fig. 2). Our code is made available at <https://github.com/mahmoodlab/Patch-GCN>.

2 Related Work

2.1 Survival Analysis in WSIs

In recent years, deep learning methods using CNNs and MIL-based approaches have been proposed for survival analysis in WSIs [13–15]. Due to the large image sizes of WSIs, many of these methods rely on selective sampling of small image ROIs for tractable training and inference, which are then used matched with patient-level outcome labels. Mobadersany *et al.* [13] proposed one of the first methods for end-to-end training with 1024×1024 image ROIs using CNNs supervised with the Partial Cox Proportional Hazard loss. Zhu *et al.* [14] developed a two-step-based approach for WSI-level survival outcome prediction, in which patches are clustered using K-Means groups using K-Means clustering method then used as inputs into a CNN. Yao *et al.* [16] similarly proposed patch-based sampling K-Means clustering to identify morphological phenotypes in WSIs.

2.2 Graph-Based Analysis in Computational Pathology

In addition to CNNs and MIL-based approaches, GCNs and other graph-based methods have received attention in computational pathology, solving problems such as cancer classification [9, 17–20], cancer grading [21–23], and survival analysis [24, 25]. Many of these approaches, however, consider only cell identities as graph nodes, which ignores important prognostic tissue features such as stroma and are confined again to small image regions [10]. In survival analysis, Chen *et al.* [25] constructed a cell-based graph for small image ROIs followed by spectral convolutions. Alternatively, Li *et al.* [24] proposed sampling patches in a WSI as nodes, followed by constructing edges between patches via feature similarity on the embedding space and using spectral convolutions. However, we argue that in using this approach for graph construction, GCNs are unable to learn context-aware features as message passing as feature interactions between adjacent image patches are not modeled.

3 Method

3.1 WSI-Graph Construction

For a given sample, let patient P , overall survival time T and censorship status C be a single triplet observation in a dataset $\{P_i, T_i, C_i\}_{i=1}^N$. In addition, let $\{W_j\}_{j=1}^K \in P$ be the set of all WSIs for P , as there may exist multiple WSIs collected for a single patient. To construct graph G for P , we first perform automatic tissue segmentation for all W_j by: 1) transforming a low-downsampled version of W_j into HSV colorspace, and then 2) using Otsu’s Binarization on the saturation channel to separate H&E-stained tissue from the background. Then, non-overlapping 256×256 instance-level image regions at $20\times$ magnification are patched and used as input for a truncated ResNet-50 model pretrained on ImageNet, which extracts a 1024-dimensional feature vector $h \in \mathbb{R}^{1024}$ via spatial average pooling after the 3rd residual block and is then packed into a node feature matrix $X_j \in \mathbb{R}^{m \times 1024}$ for M_j total patches in W_j . For each patch, we save (x,y)-coordinates from the tissue segmentation, from which we use to build an adjacency matrix A_j for each W_j via fast approximate k-NN ($k = 8$) that models a 3×3 image receptive field in CNN convolutions. Finally, we build a subgraph $G_j = (X_j, A_j)$, with the patient-level graph across all WSIs constructed as $G = \{G_j\}_{j=1}^K$ which we denote as a WSI-Graph.

In comparison to previous graph-based approaches that build neighborhoods using nearest neighbors in the embedding space, our approach is distinct in that graphs are constructed in the Euclidean space. As a result, WSI-Graphs are effectively 2D point clouds (e.g. nodes/points connected to other proximal points in a 2D planar grid), which allows us to leverage spatial convolutions that perform local neighborhood aggregation functions similar to CNNs. In comparison to CNNs, however, Path-GCN is able to tractably perform CNN-like convolution operations on thousands on extracted instance-level image features.

3.2 Patch-GCN Architecture

Message Passing: For a WSI-Graph G with M instances, we learn a differentiable function $\mathcal{F}_{\text{GCN}} : \mathbb{R}^{M \times d_{in}} \rightarrow \mathbb{R}^{M \times d_{out}}$ parameterized using a GCN that iteratively aggregates and combines node features in their spatial neighborhoods across different hidden layers via message passing. For instance, for the message passing of vertex v (that has node feature \mathbf{h}_v) with its neighboring vertices $u \in \mathcal{N}(v)$ in hidden layer $G^{(l)}$, we use the graph convolution layer $\mathcal{F}_{\text{GCN}}^{(l)}(G^{(l)}; \phi^{(l)}, \rho^{(l)}, \zeta^{(l)})$ that implement the following functions:

$$\begin{aligned} \mathbf{m}_v^{(l)} &= \rho^{(l)} \left(\left\{ \phi^{(l)} \left(\mathbf{h}_v^{(l)}, \mathbf{h}_u^{(l)}, \mathbf{h}_{e_{vu}}^{(l)} \right) \rightarrow \mathbf{m}_{vu}^{(l)} : u \in \mathcal{N}(v) \right\} \right) \\ \mathbf{h}_v^{(l+1)} &= \zeta^{(l)} \left(\mathbf{h}_v^{(l)}, \mathbf{m}_v^{(l)} \right) \end{aligned} \quad (1)$$

where $\phi^{(l)}$ is a message construction function that calculates a message $\mathbf{m}_{vu}^{(l)}$ between \mathbf{h}_v and its neighbor \mathbf{h}_u (with edge feature $\mathbf{h}_{e_{vu}}^{(l)}$), $\rho^{(l)}$ is a permutation invariant aggregation function that aggregates all messages passed to \mathbf{h}_v , and $\zeta^{(l)}$ is an update function that updates the existing node feature at v with the aggregated message $\mathbf{h}_v^{(l+1)}$. Note that the $\phi^{(l)}, \rho^{(l)}$ in Eq. 1 use similar instance-level and bag-level functions in MIL [26], in which GCN layers can be considered as performing multiple MIL operations in local graph neighborhoods, with $\zeta^{(l)}$ used as an additional differentiable function for propagating bag-level features across hidden layers in a neural network. In viewing neighborhood aggregation in GCNs has a formulation of MIL with structural neighborhood constraints, we adapt the message passing functions from DeepGCN [27] which implement $\phi^{(l)}, \rho^{(l)}, \zeta^{(l)}$ as:

$$\begin{aligned} \phi^{(l)} \left(\mathbf{h}_v^{(l)}, \mathbf{h}_u^{(l)}, \mathbf{h}_{e_{vu}}^{(l)} \right) &= \text{ReLU} \left(\mathbf{h}_u^{(l)} + \mathbf{1} \left(\mathbf{h}_{e_{vu}}^{(l)} \right) \cdot \mathbf{h}_{e_{vu}}^{(l)} \right) + \epsilon \rightarrow \mathbf{m}_{vu}^{(l)} \\ \rho^{(l)} \left(\left\{ \mathbf{m}_{vu}^{(l)} : \forall u \in \mathcal{N}(v) \right\} \right) &= \sum_{u \in \mathcal{N}(v)} \frac{\exp \left(\beta \mathbf{m}_{vu}^{(l)} \right)}{\sum_{u \in \mathcal{N}(v)} \exp \left(\beta \mathbf{m}_{vu}^{(l)} \right)} \cdot \mathbf{m}_{vu}^{(l)} \rightarrow \mathbf{m}_v^{(l)} \\ \zeta^{(l)} \left(\mathbf{h}_v^{(l)}, \mathbf{m}_v^{(l)} \right) &= \text{MLP} \left(\mathbf{h}_v^{(l)} + \mathbf{m}_v^{(l)} \right) \rightarrow \mathbf{h}_v^{(l+1)} \end{aligned} \quad (2)$$

in which $\phi^{(l)}$ is the additively combines node and edge features followed by ReLU activation, $\rho^{(l)}$ is a Softmax Aggregation scheme similar to Ilse *et al.* [28] that computes an attention weight $a_{vu}^{(l)}$ that weights how much $\mathbf{m}_{vu}^{(l)}$ should contribute to the aggregated message $\mathbf{m}_v^{(l)}$, and $\zeta^{(l)}$ additive combines the current node feature and aggregated message followed by a multilayer perceptron. Additionally, $\mathbf{1}(\cdot)$ is an indicator function when an edge feature $\mathbf{h}_{e_{vu}}^{(l)}$ exists, ϵ is a positive constant for numerical stability (set to 10^{-7}), and β is a hyperparameter for the inverse temperature in Softmax (set to 1). We argue that $\rho^{(l)}$ can be viewed as a formulation of attention pooling operation in Ilse *et al.* [28] with structural neighborhood constraints, in which attention pooling of instance-level features is performed in local graph neighborhoods instead of across the entire bag.

Learning Hierarchical Features: To learn global-level morphological features in WSIs, following [27], we make $\mathcal{F}_{\text{GCN}}^{(l)}$ a residual mapping and stack multiple layers of $\mathcal{F}_{\text{GCN}}^{(l)}$ where the output of $\mathcal{F}_{\text{GCN}}^{(l)}$ additively combines with its input.

$$G^{(l+1)} = \mathcal{F}_{\text{GCN}}^{(l)}(G^{(l)}; \phi^{(l)}, \rho^{(l)}, \zeta^{(l)}) + G^{(l)} \tag{3}$$

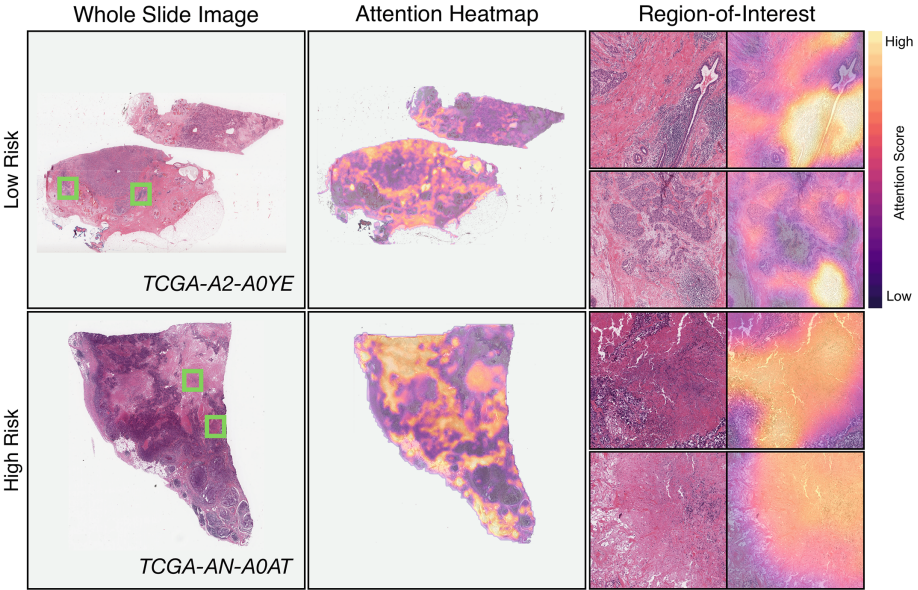


Fig. 2. Patch-GCN interpretability in BRCA survival prediction. In low risk patients, high attention regions corresponded to aggregates of lymphocytes near tumor cells, whereas in high risk patients, high attention regions corresponded to areas of tumor-associated stroma and necrosis.

We implement the spatial neighborhood aggregation backbone of Patch-GCN using $L = 4$ graph convolutional layers. As a result, each patch-based histology image feature aggregates features from other nodes in a 4-hop neighborhood, which results in an effective image receptive field size of 2302×2302 for 256×256 patches connected to its 8 nearest neighbors (Fig. 3, Supplementary Material). Furthermore, we also implement dense connections from the output of every GCN Layer to the last hidden layer of \mathcal{F}_{GCN} , so that the representation of each histology patch would be an amalgamation of its instance-level embedding and its learned surrounding context, written as $\mathbf{H}^{(L)} = [X^{(1)}, \dots, X^{(L)}]$.

Global Neighborhood Aggregation and Supervision: From the penultimate node feature matrix $\mathbf{H}^{(L)}$, following [28], we learn a global attention-based pooling layer $\mathcal{F}_{\text{AttnMIL}}(\mathbf{H}^{(L)}; \phi^{(L)}, \rho^{(L)})$ that adaptively computes a weighted sum of all node features in the graph, which generalizes aggregation function in Eq. 2 to function on all nodes in the graph, in which the node feature matrix

for the last hidden layer $\mathbf{H}^{(L)} \in \mathbb{R}^{m \times d_{\text{out}}}$ is pooled to a WSI-level embedding $\mathbf{h}_m^{(L)} \in \mathbb{R}^{1 \times d_{\text{out}}}$, which is subsequently supervised using the cross entropy-based Cox proportional loss function following [29] for survival analysis.

Implementation Details: To train Patch-GCN, we used Adam optimization with a default learning rate of 2×10^{-4} , weight decay of 1×10^{-5} , using a ResNet-50 CNN backbone pretrained on ImageNet, and trained for 20 epochs. To train with large graphs, we used 4 NVIDIA 2080 Ti GPUs with a batch size of 1 with 32 steps for gradient accumulation.

4 Experimental Setup

For this study, we used 4,370 diagnostic gigapixel WSIs across five different cancer types from The Cancer Genome Atlas: Bladder Urothelial Carcinoma (BLCA) ($n = 437$), Breast Invasive Carcinoma (BRCA) ($n = 1022$), Glioblastoma & Lower Grade Glioma (GBMLGG) ($n = 1011$), Lung Adenocarcinoma (LUAD) ($n = 515$), and Uterine Corpus Endometrial Carcinoma (UCEC) ($n = 538$). Our selection criterion in choosing these cancer types for training and evaluation were defined by: 1) dataset size, and 2) balanced distribution of uncensored-to-censored patients. On average, each WSI contained approximately $13487 \ 256 \times 256$ image patches at $20\times$ magnification, with some patients having graph sizes as large as 100000 instances.

To evaluate Patch-GCN, we trained our proposed model using 5-fold cross-validation for each cancer type, in which each dataset was split into 5 80/20 partitions for training and validation. The cross-validated concordance index (c-Index) across the validation splits was used to measure the predictive performance in correctly ranking the survival times of each patient. As qualitative assessment, we used Kaplan-Meier curves to visualize the quality of patient stratification in stratifying low and high risk patients as two different survival distributions, as well as attention-based heatmaps using the weights computed by $\mathcal{F}_{\text{AttnMIL}}$ (Fig. 2 and 4). In addition, we compared Patch-GCN against several other weakly-supervised deep learning approaches for processing in WSIs in computational pathology. As a fair comparison, we used the same survival loss function, ResNet-50 feature embeddings, and training hyperparameters in Patch-GCN.

5 Results and Discussion

5.1 Quantitative Results

In comparing our approach to other weakly-supervised learning methods for WSIs in computational pathology, Patch-GCN outperforms all prior approaches on 4 out of 5 cancer types in head-to-head comparisons, achieving an overall c-Index of **0.636** (Table 1). For cancer types such as GBMLGG which has known intertumoral and intratumoral heterogeneity, Patch-GCN achieves a c-Index of

0.824 using WSIs and shows patient stratification into distinct survival groups (Fig. 4, Supplementary Material), which empirically suggests that Patch-GCN is able to learn context-aware features via hierarchical feature aggregation in local spatial neighborhoods. In comparing Patch-GCN to permutation-invariant/MIL-based approaches, we observe that Patch-GCN improves over all methods on all 5 cancer types (9.46% performance increase over DeepAttnMISL and 3.58% performance increase over Attention MIL), which further suggests that context matters in survival outcome prediction in WSIs. In comparison to DeepGraphConv which samples random patch features from WSIs as nodes and connects these nodes on the embedding space, Patch-GCN improves on all cancer types except UCEC (2.58% performance increase), which suggests the importance of building graphs via adjacent patches rather than feature similarity in the embedding space. Though DeepGraphConv has higher c-Index on UCEC, we note that in comparison to other cancer types, cancer prognosis in UCEC correlates with global-level morphological determinants such as tumor size and depth of tumor invasion in the myometrium, rather than cell-to-cell mediated interactions between tumor cells and other cell types. BLCA is a similar cancer type to UCEC that also depends on the depth of invasion into the bladder wall, but because the bladder wall is thinner than the myometrium, the invasion may be adequately captured via a limited receptive field, hence better Patch-GCN performance on that cancer type.

Table 1. c-Index performance comparisons of Patch-GCN against prior state-of-the-art weakly-supervised approaches on 5 cancer types in the TCGA.

Models	BLCA	BRCA	GBMLGG
MIL (Deep Sets) [26]	0.500 ± 0.000	0.500 ± 0.000	0.498 ± 0.014
Attention MIL [28]	0.536 ± 0.038	0.564 ± 0.050	0.787 ± 0.028
DeepAttnMISL [16]	0.504 ± 0.042	0.524 ± 0.043	0.734 ± 0.029
DeepGraphConv [24]	0.499 ± 0.057	0.574 ± 0.044	0.816 ± 0.025
Patch-GCN (Ours)	0.560 ± 0.034	0.580 ± 0.025	0.824 ± 0.024
	LUAD	UCEC	Overall
MIL (Deep Sets) [26]	0.496 ± 0.008	0.500 ± 0.000	0.499
Attention MIL [28]	0.559 ± 0.060	0.625 ± 0.057	0.614
DeepAttnMISL [16]	0.548 ± 0.050	0.597 ± 0.059	0.581
DeepGraphConv [24]	0.552 ± 0.058	0.659 ± 0.056	0.620
Patch-GCN (Ours)	0.585 ± 0.012	0.629 ± 0.052	0.636

5.2 Attention Visualization

To understand how Patch-GCN uses morphological features to predict risk, we visualized heatmaps using the attention weights from the attention pooling layer and utilized two trained pathologists to assess high-attention image regions. Across all cancers, we observed that in high risk patients, the network assigned

high attention to necrosis, dense tumor aggregates, and regions of desmoplastic stroma containing tumor infiltrates, which are indicative of tumor invasion and proliferation (Fig. 2). In low risk patients, we observe that lymphocyte aggregates and normal stroma were frequently assigned high attention, which corroborates with the prognostic significance of stroma [10]. Figure 2 shows exemplar low and high risk cases in BRCA, with lymphocytes adjacent to tumor cells and infiltrating normal stroma given high attention in low risk patients, while necrosis and desmoplastic stroma were given high attention in high risk patients.

5.3 Conclusion

Despite the progress made in weakly-supervised deep learning in computational pathology, many current approaches are not context-aware in modeling important local- and global-level morphological features in the tumor microenvironment. In this work, we present Patch-GCN, a context-aware, attention-based graph convolutional network for survival analysis using WSIs. In comparing Patch-GCN to permutation-invariant network architectures that learn only instance-level morphological features, we observe that Patch-GCN outperforms all prior approaches on 5 cancer types in the TCGA. Moreover, we demonstrate the improvement in connecting nodes via adjacent image patches, which allows node aggregation in GCNs to learn such coarse-grained to fine-grained topological structures in the tumor microenvironment. Our approach is adaptable to any weakly-supervised learning task in computational pathology that uses slide-level or patient-level labels, and contributes towards a more holistic view of representation learning in the tumor microenvironment.

Acknowledgements. Funding: This work was supported in part by internal funds from BWH Pathology, Google Cloud Research Grant, Nvidia GPU Grant Program, and NIGMS R35GM138216 (F.M.). R.J.C. was additionally supported by the NSF Graduate Fellowship. The content is solely the responsibility of the authors and does not reflect the official views of the National Institutes of Health, National Institute of General Medical Sciences or the National Science Foundation.

References

1. Yu, K.H., et al.: Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat. Commun.* **7**(1), 1–10 (2016)
2. Campanella, G., et al.: Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* **25**(8), 1301–1309 (2019)
3. Courtiol, P., et al.: Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nat. Med.* **25**(10), 1519–1525 (2019)
4. Wulczyn, E., et al.: Interpretable survival prediction for colorectal cancer using deep learning. *NPJ Digit. Med.* **4**(1), 1–13 (2021)
5. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data efficient and weakly supervised computational pathology on whole slide images. *Nat. Biomed. Eng.* **5**, 555–570 (2020)

6. Bandi, P., et al.: From detection of individual metastases to classification of lymph node status at the patient level: the CAMELYON17 challenge. *IEEE Trans. Med. Imaging* **38**(2), 550–560 (2018)
7. Balkwill, F.R., Capasso, M., Hagemann, T.: The tumor microenvironment at a glance. *J. Cell Sci.* **125**(23), 5591–5596 (2012)
8. Saltz, J., et al.: Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep.* **23**(1), 181–193 (2018)
9. Zhao, Y., et al.: Predicting lymph node metastasis using histopathological images based on multiple instance learning with deep graph convolution. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4837–4846 (2020)
10. Beck, A.H., et al.: Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci. Transl. Med.* **3**(108), 108ra113 (2011)
11. Abdul Jabbar, K., et al.: Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. *Nat. Med.* **26**(7), 1054–1062 (2020)
12. Shaban, M., et al.: A novel digital score for abundance of tumour infiltrating lymphocytes predicts disease free survival in oral squamous cell carcinoma. *Sci. Rep.* **9**(1), 1–13 (2019)
13. Mobadersany, P., et al.: Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc. Natl. Acad. Sci.* **115**(13), E2970–E2979 (2018)
14. Zhu, X., Yao, J., Zhu, F., Huang, J.: WSISA: making survival prediction from whole slide histopathological images. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7234–7242 (2017)
15. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat. Biomed. Eng.* **5**(6), 555–570 (2021)
16. Yao, J., Zhu, X., Jonnagaddala, J., Hawkins, N., Huang, J.: Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. *Med. Image Anal.* **65**, 101789 (2020)
17. Anand, D., Gadiya, S., Sethi, A.: Histograms: graphs in histopathology. In: *Medical Imaging 2020: Digital Pathology*, vol. 11320, p. 113200O. International Society for Optics and Photonics (2020)
18. 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
19. 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
20. Pati, P., et al.: Hierarchical cell-to-tissue graph representations for breast cancer subtyping in digital pathology. *arXiv e-prints* [arXiv-2102](https://arxiv.org/abs/2102.00000) (2021)
21. Zhou, Y., Graham, S., Alemi Koohbanani, N., Shaban, M., Heng, P.A., Rajpoot, N.: CGC-Net: cell graph convolutional network for grading of colorectal cancer histology images. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops* (2019)
22. Wang, J., Chen, R.J., Lu, M.Y., Baras, A., Mahmood, F.: Weakly supervised prostate TMA classification via graph convolutional networks. In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pp. 239–243. IEEE (2020)

23. Javed, S., Mahmood, A., Werghi, N., Benes, K., Rajpoot, N.: Multiplex cellular communities in multi-gigapixel colorectal cancer histology images for tissue phenotyping. *IEEE Trans. Image Process.* **29**, 9204–9219 (2020)
24. Li, R., Yao, J., Zhu, X., Li, Y., Huang, J.: Graph CNN for survival analysis on whole slide pathological images. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) *MICCAI 2018*. LNCS, vol. 11071, pp. 174–182. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_20
25. Chen, R.J., et al.: Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Trans. Med. Imaging* (2020)
26. Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R., Smola, A.: Deep sets. In: *Advances in Neural Information Processing Systems (NeurIPS)* (2017)
27. Li, G., Muller, M., Thabet, A., Ghanem, B.: DeepGCNs: can GCNs go as deep as CNNs? In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9267–9276 (2019)
28. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: *Proceedings of the 35th International Conference on Machine Learning (ICML)*, pp. 2132–2141 (2018)
29. Zadeh, S.G., Schmid, M.: Bias in cross-entropy-based training of deep survival networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**, 3126–3137 (2020)