Feature-Enhanced Graph Networks for Genetic Mutational Prediction Using Histopathological Images in Colon Cancer

Kexin Ding[1], Qiao Liu[2], Edward Lee[3], Mu Zhou[4], Aidong Lu[1], and Shaoting Zhang[5(✉)]

[1] Department of Computer Science, UNC Charlotte, Charlotte, NC, USA
[2] Department of Automation, Tsinghua University, Beijing, China
[3] Department of Radiology, Stanford University, Stanford, CA, USA
[4] SenseBrain Research, Santa Clara, CA, USA
[5] SenseTime Research, Shanghai, China
szhang16@uncc.edu

Abstract. Mining histopathological and genetic data provides a unique avenue to deepen our understanding of cancer biology. However, extensive cancer heterogeneity across image- and molecular-scales poses technical challenges for feature extraction and outcome prediction. In this study, we propose a feature-enhanced graph network (FENet) for genetic mutation prediction using histopathological images in colon cancer. Unlike conventional approaches analyzing patch-based feature alone without considering their spatial connectivity, we seek to link and explore non-isomorphic topological structures in histopathological images. Our FENet incorporates feature enhancement in convolutional graph neural networks to aggregate discriminative features for capturing gene mutation status. Specifically, our approach could identify both local patch feature information and global topological structure in histopathological images simultaneously. Furthermore, we introduced an ensemble strategy by constructing multiple subgraphs to boost the prediction performance. Extensive experiments on the TCGA-COAD and TCGA-READ cohort including both histopathological images and three key genes' mutation profiles (APC, KRAS, and TP53) demonstrated the superiority of FENet for key mutational outcome prediction in colon cancer.

Keywords: Histopathological image analysis · Graph convolutional networks · Gene mutation prediction

1 Introduction

Colon cancer [1,2] is the third common cancer worldwide that accounts for 13% of all new cancer incidence and approximately 8% of all cancer deaths. Especially

This study has been partially supported by fund of STCSM (19511121400).

A. L. Martel et al. (Eds.): MICCAI 2020, LNCS 12262, pp. 294–304, 2020.
https://doi.org/10.1007/978-3-030-59713-9_29

colon adenoma and carcinoma are known to occur through a series of histopathological changes due to key genetic alterations [3–5]. Therefore, the prediction of genetic mutations over the course of cancer evolution is highly desired towards accurate detection and diagnosis of colon cancer.

Mining histopathological and genetic profiles provides a unique avenue to deepen our understanding of cancer biology across scales. Particularly, genetic mutations play key roles in colon cancer evolution at all clinical stages (Fig. 1). For example, KRAS gene mutation has been proven to be an independent prognostic factor in patient with advanced colon cancer [5]. Also, APC gene mutation triggers chain of molecular and histological changes, leading to increased growth of colon cancer cells [6]. In parallel, whole slide images (WSIs) permit extensive cell-level characterization of individual patients. Thus increasing emphasis has been placed on extracting quantitative features from WSIs for outcome assessment [7,8]. Preliminary evidences suggested the usefulness of quantitative features extracted from large-scale image patches [9–13]. However, conventional approaches were merely focused on image patches and thus unable to consider rich topological structures as shown in WSIs. Especially local and spatial connectivity of image-based findings, critical to understand characteristics of cancer heterogeneity [14], has not been explicitly analyzed.

To explore the topological structure of WSIs, we introduce a graph-based analysis with a goal to capture both spatial and local histopatholgical variations. Specifically, we focus on analyzing spatial-based graph convolutional networks (Spatial-GCNs) due to their advances in network flexibility [15–17]. For example, spatial-GCNs allows convolutional operations locally on each node with weight sharing across locations and structures [18]. Thus it is more convenient to integrate node features with their neighborhood information in spatial-GCNs, compared with Spectral-based graph convolutional networks (Spectral-GCNs) that commonly requires entire graph Laplacian embedding. In addition, spatial-GCNs offers mechanisms of aggregators for feature integration. For example, GraphSage [15], GCN [19] and GIN [17] models demonstrate their learning ability using max-, mean-, or sum-pooling aggregators. However, it remains uncertain about how can we implement an efficient graph structure to characterize histopathological images, especially for the classification of non-isomorphic topological structures in our study.

In this study, we proposed the **f**eature-**e**nhanced graph **net**work (FENet), as a novel graph-driven approach for gene mutation prediction in patients with colon cancer (Fig. 1). To tackle the challenge of cancer heterogeneity, our FENet model consists of multiple subgraphs construction from histopathological image patches. Conceptually, FENet can be considered as an ensemble learning solution for predicting genetic signal variations with enhanced feature integration. Our major contributions are summarized as follows.

– We introduced an efficient transformation method between WSIs and graph structured data. Focusing on generating spatially-connected graphs, our approach can link and explore local feature and global topological structure of WSIs simultaneously.

– We developed a feature-enhanced model to underscore discriminative feature learning. In this architecture, we improved the ability of distinguishing non-isomorphic topological structure, and adaptively selected the node representations from the different graph convolution layers.
– Our ensemble strategy of network models alleviates cancer heterogeneity so that integration of multiple subgraph outcomes leads to a significant improvement of prediction performance.

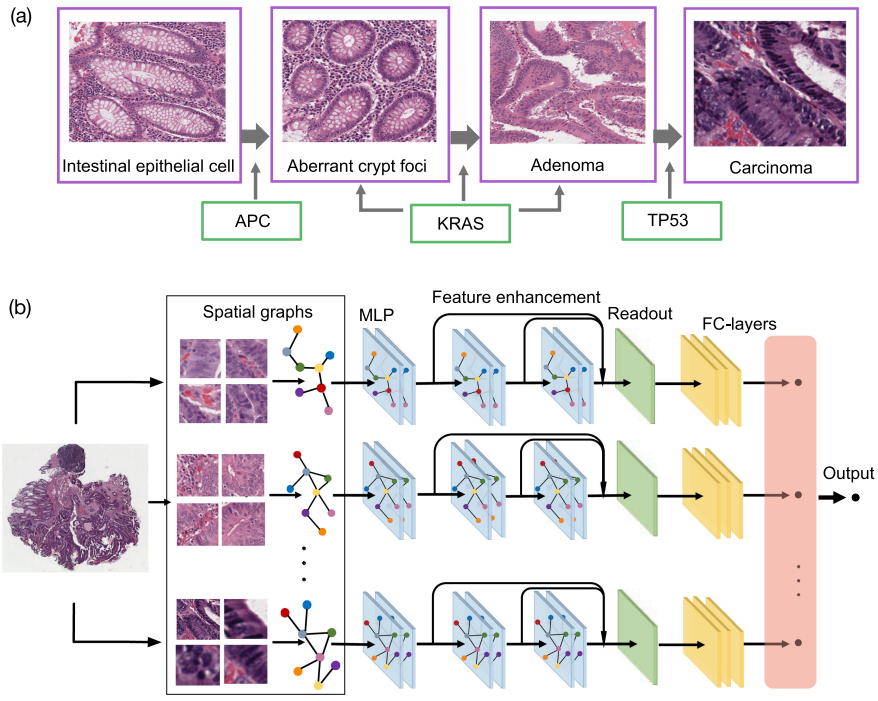

Fig. 1. (a) Illustration of colon cancer evolution in histopathology over time and its key genetic mutations. (b) the proposed FENet networks architecture.

2 Methodology

Overview of FENet Model. Our FENet is constructed by multiple subgraphs using image patches from each patient's WSI. We particularly underscored the spatially-connected subgraph construction strategy that was seldom addressed in prior studies [9–13]. To increase feature learning performance, we then described our feature-enhanced mechanism to aggregate the features of neighboring patches and combine the aggregated feature as the updated central

node representation. Moreover, our feature-enhanced mechanism could adaptively select the node representations from different graph convolution layers. Finally, an ensemble strategy was introduced to combine the prediction results of subgraphs for predicting important mutation statuses (i.e., mutated and non-mutated classes).

Spatially-Connected Subgraph Construction. For each whole slide image (WSI), we randomly selected a set of sampled patches $P = \{P_1, P_2, \ldots, P_N\}$ from all patches generated from WSIs (N is the number of patches). Random selection always maintains the distribution of the original feature underlying WSI, which provides a comprehensive description of WSIs. In our study, we defined patches as graph nodes in each subgraph, and the spatial distance between two patches determines whether there exists a graph edge. We emphasized the analysis of non-isomorphic graphs because informative image patches caused by cancer mutation can be always in various spatial locations in WSIs [14]. This property is opposed to isomorphic graphs which strictly share the same adjacency neighborhood matrix. Therefore, we first used a pre-trained ResNet18 model to extract high-level features (node attribute) of all individual patches within a subgraph. We constructed graph representation for each subgraph by calculating its adjacent matrix with criterion that the edge exists if the spatial distance is below a fixed threshold. Precisely, we measured the spatial distance by directly calculating the Euclidean distance of two patches mapped on the original WSI. Algorithm 1 describes the details of the subgraph construction.

Algorithm 1. Spatially-connected Subgraph Construction Algorithm

Input: A set of selected patches in one slide $P = P_1, P_2, ..., P_N$; a set of top-left corresponding coordinates of the selected patches $C_{p_i} = (x_1, y_1), (x_1, y_1), ..., (x_N, y_N)$; **N** is the number of patches.
Threshold **t** is the distance

Output: Adjacency Matrix **A** and Feature Matrix **X**

function CONSTRUCTING(A, P, C_{p_i}, N, t)
  $X \leftarrow$ Feature_extractor(P)
  for $i \leftarrow 1$ to N **do**
    $(x_j, y_j) \leftarrow c_{p_j}$
    $s_{ij} \leftarrow \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$
    if $s_{ij} < t$ **then**
      $A_{ij} \leftarrow 1$
    end if
  end for
  return A,X
end function

Feature-Enhanced Mechanism. Spatial-based convolutional networks utilize a neighborhood aggregation strategy that iteratively updates the representation of a central node by AGGREGATE and COMBINE [15,16]. The AGGREGATE can aggregate neighboring node representations of the center node, while the COMBINE combines the neighborhood node representation with the center node representation to obtain the updated center node representation. After k iterations of aggregation and combination, a node's representation captures the structural information within its k-hop network neighborhood. Formally, the k^{th} layer spatial-based graph convolutional network can be represented as

$$a_v^{(k)} = \text{AGGREGATE}(\{h_u^{(k-1)} : u \in \mathcal{N}(v)\}) \tag{1}$$

$$h_v^{(k)} = \text{COMBINE}(h_{v-1}^{(k)}, a_v^{(k)}) \tag{2}$$

Where $h_v^{(k)}$ is the feature vector of node v at the k^{th} layer. $\mathcal{N}(v)$ is a set of nodes adjacent to v. As illustrated in [17], the selection of AGGREGATE and COMBINE is critical to capture the graph topological structures. The traditional aggregation strategy using max-pooling (GraphSage) considers multiple node representations as one node representation while ignoring multiplicities. Alternatively, the mean-pooling (GCN) captures the statistical and distributional information within the graph rather than the exaction of the entire graph structure. These measurements can be useful if the graphs are isomorphic (i.e., with a strictly same adjacency matrix). However, the topological structure of subgraphs are non-isomorphisms in our study, improving the ability to distinguish topological non-isomorphism is a critical issue. As proved in [17], sum-pooling captures the full structural information of the entire graph representation. Additionally, we use multi-layer perceptrons (MLPs) with Rectified Linear Unit (ReLU) and batch normalization [20] to model in each graph neural network layers. Therefore, FENet updates node feature representation $h_v^{(k)}$ as:

$$h_v^{(k)} = \text{MLP}^{(k)}(h_{v-1}^{(k)} + \sum_{u \in \mathcal{N}(v)} h_u^{(k-1)}) \tag{3}$$

It is known that nodes in the central and boundary regions of a graph requiring different frequencies of aggregation to achieve optimal performance [21]. Our WSI-based graph contains both central and marginal nodes since each WSI has a unique spatial distribution of cancerous regions. To consider all topological structural information, we emphasized the useful information from all depths of network layers. Therefore, we aggregated node representation from each previous layer to the last layer. By this design, the network could adaptively select the most meaningful information during the training process and find the desired representation for each node. To achieve the subgraph classification, we introduced the READOUT function to converts node representations into a graph representation. To guarantee that each part in the network is injective, we also selected GlobalAddPooling as a READOUT function to aggregate information from all nodes and to convert it to the entire graph representation. Formally,

FENet aggregates layers information and converts node representation to graph representation h_G by:

$$h_G = \text{READOUT}(h_v^1 \| h_v^2 \| ... \| h_v^k, v \in G) \tag{4}$$

where $\|$ is the feature concatenation. The feature concatenation could aggregate node representation in different graph convolutional layers. After READOUT representation, the final prediction outcome for each subgraph is classified by fully-connected layers.

Ensemble Strategy. Our ensemble strategy utilized the majority vote to aggregate all subgraphs' prediction outcomes derived from the same WSI. The ensemble strategy is motivated by two major findings. First, since a WSI can contain a large number of patches (e.g., 10k) that allows us to explore the diversity of WSI characteristics via individual subgraphs. Second, analyzing individual patches creates a sizeable computational burden as shown in conventional CNN-based approaches. The ensemble strategy, therefore, achieves a good trade-off between informative representation of WSI and computational burden. We highlighted that ensemble learning allowed FENet to increase its generalization power by exploiting the advance of multiple spatial subgraphs of which the predictive error can be reduced by majority vote.

3 Experiments and Results

Dataset. We collected whole slide images (WSIs) from The Cancer Genome Atlas Colon Adenocarcinoma (TCGA-COAD) dataset [22], which contains 421 WSIs with colon tumors. We identified the associated colon cancer genetic mutational profiles from Cbioportal [23]. Data exclusion criteria included that we removed 40 WSIs with noisy stained annotation. We also removed patient data without key mutational profiles. We finally analyzed the total number of 274 patients' WSIs with a resolution of 40X (0.25 microns/pixel). For each type of mutational profile in each patient, we assigned the outcome label as a positive class if mutated and as a negative class if non-mutated. We focus on three key genes (APC, KRAS, and TP53) that are significantly associated with colon cancer evolution over various clinical stages with treatment impact [22]. We found that 70% of samples contain APC mutation, 60% of samples contain TP53 mutation, and 40% of samples contain KRAS mutation. We also collected 30 WSIs from The Cancer Genome Atlas Rectum Adenocarcinoma (TCGA-READ) as an external dataset. We found that 75% of samples contain APC mutation, 70% of samples contain TP53 mutation, and 36% of samples contain KRAS mutation.

Experimental Settings and Implementations. In data preprocessing, all slides are color-normalized by Macenko's method [24], and the foreground is segmented using OTSU [25]. To obtain a tumor region in WSI, we trained a

tumor detection model by performing a pre-trained ResNet18 on the NCT-CRC-HE-100K dataset [26]. A fault-tolerant tumor region delineation is acceptable due to our patch-focused analysis. We then generated non-overlapping patches with a size of 224 * 224 that was resized from raw 512 * 512 patches within the tumor region. We then randomly generated five different subsets of patches to build subgraphs. The number of selected patches in each subset was set to 1,000 due to the consideration of computational efficiency [27].

For graph construction, we used a pre-trained ResNet18 to extract features from patches as the feature matrix. For the adjacency matrix, we calculated the Euclidean distance between patches' coordinate values recorded in their raw WSI and determine the connection by a fixed spatial distance threshold. Precisely, we calculated the mode value from the statistical distribution of the spatial Euclidean distances among all pairs of patches to determine the threshold value. Finally, the feature matrix and the adjacency matrix were used for the non-isomorphic graph representation of WSI.

In the experiments, we evaluated our FENet with multiple competing methods on TCGA-COAD cohort. The performances are achieved from 10-fold cross-validation and reported the mean accuracy and AUC among 10 times experiments for each prediction task. We guarantee patches in the same WSI will not appear in different sets simultaneously. Moreover, to show the generalization power, we trained FENet on TCGA-COAD and tested it on TCGA-READ. To facilitate model training, we augmented patch samples from the minority class in the training set to balance the number of positive and negative samples. Notably, we always kept the real positive and negative ratios at the testing stage. In the training set, we dropped a fraction of edges (drop rate = 0.3) to reduce potential overfitting as commonly done in [28]. The optimal hyperparameters were obtained by a grid search. For a fair comparison, we always used the same hyperparameters setting in all experiments to ensure differences only come from the variants of the methods. All models were trained with initial learning rate 1e−3, batch size 64 by Adam optimizer [29] with a weight decay 5e−3 and the cross-entropy loss.

Method Comparison and Ablation Study. We compared our approach with multiple state-of-the-art models. GCN [19] is a spectral-GCNs with mean-pooling aggregation. GraphSAGE [15] is a spatial-GCNs model applying node embedding with max-pooling aggregation. GAT [18] is a spatial-GCNs model that leverages masked self-attentional layers to graph convolutions. We used VGG16 [31] and ResNet18 [32] as a baseline CNN-based model. Additionally, we designed several ablation studies. First, we introduced the Perceptron-3 model by replacing model's MLPs by 1-layer perceptrons with sum-pooling as the aggregator. Meanwhile, Perceptron-3 comes without ensemble strategy and the operation feature concatenation (i.e., aggregating node representation from each layer to the last layer). Second, MLP-n model removes both the ensemble strategy and feature concatenation in the FENet-n, where n is the number of MLP in the model. Third, the FENet-n w/o ensemble model keeps all components in the

FENet (e.g., feature concatenation, MLPs and sum-pooling) except the ensemble strategy. Finally, the FENet-3 is our proposed approach with full components.

Table 1. The prediction performance of mean accuracy and variance of FENet with competing approaches and ablation studies.

Network architecture	Network type	TP53	KRAS	APC
VGG16 [30]	Deep CNN	59.89 ± 0.84	58.28 ± 1.12	56.36 ± 0.83
ResNet18 [31]	Deep CNN	62.49 ± 2.22	61.33 ± 2.51	70.86 ± 0.18
GCN [19]	Spectral-based	60.98 ± 0.71	62.21 ± 0.89	70.85 ± 0.24
ChebNet [32]	Spectral-based	60.47 ± 0.86	61.62 ± 0.93	68.70 ± 0.76
GraphSAGE [15]	Spatial-based	60.31 ± 0.95	63.98 ± 1.27	71.39 ± 1.14
GAT [16]	Spatial-based	62.47 ± 0.86	60.47 ± 1.42	68.79 ± 0.92
Perceptron-3	Ablation analysis	64.50 ± 0.25	65.75 ± 0.62	68.89 ± 0.68
MLP-2	Ablation analysis	66.03 ± 0.36	64.21 ± 0.29	70.37 ± 0.25
FENet-2 w/o essemble	Ablation analysis	65.45 ± 0.56	67.57 ± 0.52	71.72 ± 0.22
MLP-3	Ablation analysis	65.63 ± 0.14	66.29 ± 0.08	71.77 ± 0.21
FENet-3 w/o essemble	Ablation analysis	67.44 ± 0.36	69.87 ± 0.25	73.18 ± 0.31
FENet-3	Our method	$\mathbf{77.00 \pm 0.01}$	$\mathbf{80.38 \pm 0.01}$	$\mathbf{79.93 \pm 0.01}$

Results and Analysis. In Table 1, our FENet demonstrated leading performance across all three key mutational prediction tasks. Even viewing our perceptron-3 model without ensemble learning or feature concatenation, it is remarkable that Perceptron-3 was quite competitive to the graph-based baseline models (e.g., GCN, GraphSAGE and GAT). This key finding can be ascribed to the contribution of our design of sum-pooling aggregator followed by [17], leading to improved ability to distinguish different non-isomorphic topological structures. Notably, our task is challenging because traditional CNN-based approaches working on patches alone may not be able to perform well (e.g., VGG16). In Fig. 2, we observed that the feature concatenation performed strongly for improving the performance by comparing MLP-3 model and FENet-3 without ensemble. A similar finding was confirmed by viewing MLP-2 model and FENet-2 without ensemble. Furthermore, we recognized the boosted performance with our ensemble strategy on all three prediction tasks (Fig. 2). Overall, the ensembled strategy greatly increased the generalization power of our FENet probably due to its ability to capture diverse topological information of WSI derived from our multiple spatially-connected subgraphs. Besides the superior performance on TCGA-COAD cohort, the proposed model also showed a relatively good performance on TCGA-READ. The test accuracy of TP53 on TCGA-READ is 0.7325, and AUC is 0.7522. Both test accuracy and AUC of KRAS and APC are higher than 0.6. The comparative performance on TCGA-READ shows a potential generalization power of the proposed model.

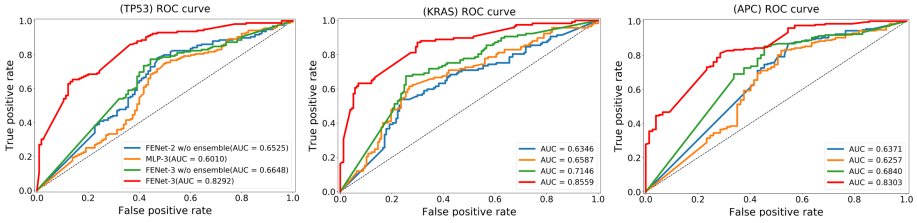

Fig. 2. The ROC-curve comparison of our approach with or without ensemble strategy. FENet with ensemble strategy brings the boosted performance over other methods.

4 Conclusion

In this study, we have proposed the FENet to tackle the challenging problem of predicting gene mutational status in colon cancer using histopathological images. Our findings supported that the convergence of multi-scale cancer data leads to novel insights into modeling cancer biology. To address the cancer heterogeneity, we highlighted the importance of exploring spatial and local connections of image patches via graph construction, which offers an efficient means for important molecular outcome prediction. Extensive experiments suggested that our model with a subgraph ensemble strategy outperformed current state-of-the-art approaches. In the future work, we plan to perform and validate a large-scale of our analysis across different types of cancers and thus gain more insights into multi-scale cancer data integration.

References

1. Mármol, I., Sánchez-de-Diego, C., Pradilla, D.A.: Colorectal carcinoma: a general overview and future perspectives in colorectal cancer. Int. J. Mol. Sci. **18**, 197 (2017)
2. Bray, F.: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. **68**, 394–424 (2018)
3. Iacopetta, B.: TP53 mutation in colorectal cancer. In: Human Mutation (2003)
4. Armaghany, T., Wilson, J.D., Chu, Q., Mills, G.: Genetic alterations in colorectal cancer. Gastrointest. Cancer Res. **5**, 19–27 (2012)
5. Jancik, S., Drabek, J., Radzioch, D., Hajduch, M.: Clinical relevance of KRAS in human cancers. J. Biomed. Biotechnol. **2010**, 150960 (2010)
6. Fodde, R.: The APC gene in colorectal cancer. Eur. J. Cancer **38**(7), 867–71 (2002)
7. Kather, J.N.: Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat. Med. **25**, 1054–1056 (2019)
8. Coudray, N., et al.: Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat. Med. **24**(10), 1559–1567 (2018)
9. Li, Z., Zhang, X., Muller, H., Zhang, S.: Large-scale retrieval for medical image analytics: a comprehensive review. Med. Image Anal. **43**, 66–84 (2018)
10. Ghaznavi, F., Evans, A., Madabhushi, A., Feldman, M.: Digital imaging in pathology: whole-slide imaging and beyond. Annu. Rev. Pathol. **8**, 331–359 (2013)

11. Mobadersany, P., et al.: Predicting cancer outcomes from histology and genomics using convolutional networks. Proc. Nat. Acad. Sci. U.S.A. **115**, E2970–E2979 (2018)
12. Zhang, X., Su, H., Yang, L., Zhang, S.: Fine-grained histopathological image analysis via robust segmentation and large-scale retrieval. In: IEEE Conference on Computer Vision and Pattern Recognition (2015)
13. Duan, Q., et al.: SenseCare: a research platform for medical image informatics and interactive 3D visualization. https://arxiv.org/abs/2004.07031
14. Heindl, A., Nawaz, S., Yuan, Y.: Mapping spatial heterogeneity in the tumor microenvironment: a new era for digital pathology. Lab. Investig. **95**, 377–384 (2015)
15. Hamilton, W.L., Ying, R., Leskovec, J.: Inductive representation learning on large graphs. In: Neural Information Processing Systems (2017)
16. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. In International Conference on Learning Representations (2018)
17. Xu, K, Hu, W., Leskovec, J., Jegelka, S.: How powerful are graph neural networks. In: International Conference on Learning Representations (2019)
18. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Yu, P.S.: A comprehensive survey on graph
19. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (2016)
20. Ioffe, S., et al.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning (2015)
21. Xu, K., Li, C., Tian, Y., Sonobe, T., Kawarabayashi, K., Jegelka, S.: Representation learning on graphs with jumping knowledge networks. In: International Conference on Machine Learning (2018)
22. Kirk, S., et al.: Radiology data from the cancer genome atlas colon adenocarcinoma [TCGA-COAD] collection. In: The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2016.HJJHBOXZ
23. Cbioportal Homepage. https://www.cbioportal.org/
24. Macenko, M., et al.: A method for normalizing histology slides for quantitative analysis. In: Proceedings of IEEE International Symposium on Biomedical Imaging (2011)
25. Otsu, N.: A threshold selection method from gray-level histogram. IEEE Trans. Syst. Man Cybern. B Cybern. **9**, 62–66 (1979)
26. Kather, J.N., Halama, N, Marx, A.: 100,000 histological images of human colorectal cancer and healthy tissue (Version v0.1) [Data set]. Zenodo (2018). https://doi.org/10.5281/zenodo.1214456
27. Li, R., Yao, J., Zhu, X., Li, Y., Huang, J.: Graph CNN for survival analysis on whole slide pathological images. In: Medical Image Computing and Computer Assisted Intervention (2018)
28. Rong, Y., Huang, W., Xu, T., Huang, J.: DropEdge: towards deep graph convolutional networks on node classification. In: International Conference on Learning Representations (2020)
29. Kingma, D., Jimmy B.: Adam: a method for stochastic optimization. In: International Conference on Learning Representations (2015)
30. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations (2015)

31. He, K., Zhang, X., Ren, S., and Sun, J. : Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (2016)
32. Defferrard, M., Bresson, X., Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering. In: Neural Information Processing Systems (2016)