

Capturing Cellular Topology in Multi-Gigapixel Pathology Images

Wenqi Lu, Simon Graham, Mohsin Bilal, Nasir Rajpoot, Fayyaz Minhas

Department of Computer Science, University of Warwick, UK

wenqi.lu.1@warwick.ac.uk

Abstract

*In computational pathology, multi-gigapixel whole slide images (WSIs) are typically divided into small patches because of their extremely large size and memory requirements. However, following this strategy, one risks losing visual context which is very important in the development of machine learning models aimed at diagnostic and prognostic assessment of WSIs. In this paper, we propose a novel graph convolutional neural network based model (called **Slide Graph**) which overcomes these limitations by building a graph representation of the cellular architecture in an entire WSI in a bottom-up manner. We evaluate Slide Graph for prediction of the status of human epidermal growth factor receptor 2 (HER2) and progesterone receptor (PR) expression from WSIs of H&E stained tissue slides of breast cancer. We demonstrate that the proposed model outperforms previous state-of-the-art methods and is more computationally efficient. The proposed paradigm of WSI-level graphs can potentially be applied to other problems in computational pathology as well.*

1. Introduction

Histopathology slide examination under the microscope by a pathologist is considered the *gold standard* for cancer diagnosis and determining treatment options for a patient. In recent years, digital pathology has revolutionized the clinical practice of pathology through the use of slide scanners to create multi-resolution multi-gigapixel whole slide images (WSIs). Increasing adoption of WSIs for routine diagnosis has also given rise to the new discipline of computational pathology, which aims at the development of machine learning and artificial intelligence techniques to automatically analyse WSIs [1, 2, 3]. These tools can improve pathology workflow by providing a more objective and reproducible results, leading to better patient care [4, 5].

A major limitation of existing methods in computational pathology [6, 7, 8] stems from the computational complexity associated with whole slide images. Training a deep learning network on entire WSIs at full resolution is compu-

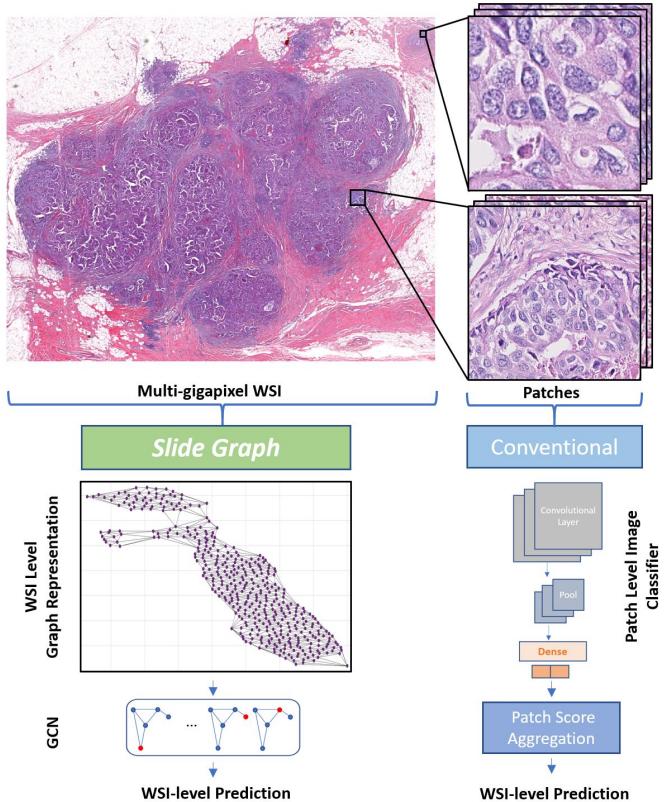


Figure 1. The proposed *Slide Graph* Model vs. conventional methods.

tationally intractable as the size of WSIs at the highest resolution can be up to $150,000 \times 100,000$ pixels. Typically, a two-step approach is used to deal with WSIs of such a large size (Fig. 1) [9, 10]. First, the image is sub-divided into small patches, where each patch is processed independently in the neural network [8]. Then, the predicted scores for each patch within a WSI are aggregated, usually by combining (pooling) their results with various aggregating strategies such as average pooling, max pooling¹ and majority

¹Here, pooling refers to the aggregation of multiple predictions, not down-sampling

voting [11, 12, 13].

The use of patch-level analysis leads to two major problems. First, patches provide a limited *visual context*. The optimal resolution and patch size for analysis are highly problem-dependent [7]. Under a specific patch size, patches drawn at a high magnification level lead to less contextual and spatial information whereas patches at lower magnification levels may not capture cell-level features (see patches in Fig. 1). Consequently, a patch-level machine learning method cannot capture the overall organization and structure of the tissue in a WSI. Secondly, in most prediction problems in computational pathology, only WSI-level labels are often available and it is non-trivial to model the association of different patches with a target class. Weakly supervised machine learning methods such as Multiple Instance Learning (MIL) have been proposed to alleviate incomplete knowledge about labels of training patches and aggregates patch level predictions into WSI level classification. [14, 15, 16, 17]. However, these methods are unable to model the geometric structure of the tissue at both global and local levels.

The cell-graph technique [18, 19, 20] was introduced to learn the structure-function relationship by modeling geometric structure of the tissue using graph theory. It is based on the assumption that cells in a tissue can organize in a certain way for specific functional states, such as receptor status in computational pathology. Such cell-graph can have different types, such as Delaunay Triangles [21, 22], Voronoi Diagrams, Minimum Spanning Trees (MST), and Cell Cluster Graphs (CCG) [23]. Yener in [24] explored various cell-graph constructions to establish a quantitative relationship between the geometric structure and functional states. Cell-graph constructions have been successfully used to characterize spatial proximity of histopathologic primitives in tasks, such as survival prediction in lung cancer [25], risk categories prediction in breast cancer [26], distant metastasis prediction in colorectal cancer [27]. However those graph-based methods with deep learning classifiers were all trained on a per-patch basis which have limited visual context. Extra patch-based voting methods are necessary to assess the functional state of a given WSI.

In this paper, we propose a graph-based model called *Slide Graph* to handle these limitations of existing methods. Instead of extracting small patches from the WSI and doing analysis on a limited visual field for prediction, we introduce a pipeline which constructs a graph from the nuclei-level to the entire WSI-level (Fig. 1). A graph convolutional neural network is then used for WSI-level prediction. This method accounts for both cell-level information and contextual information by modelling cellular architecture and interactions in the form of a graph. We demonstrate the effectiveness of the proposed scheme on two clinically relevant prediction problems in computational pathology: prediction

of the status of human epidermal growth factor receptor 2 (HER2) and progesterone receptor (PR) from breast cancer hematoxylin and eosin (H&E) stained whole slide images.

Overall, our main contributions in this paper can be summarized as follows:

- To the best of our knowledge, Slide Graph is the first method which can generate slide-level predictions by using a graph representation of the cellular interconnection geometry in a whole slide image.
- Slide Graph makes use of nuclei type and region features to represent the complex organization of cells and the overall tissue micro-architecture. The proposed network outperforms the state-of-the-art methods by a significant margin in HER2 and PR status prediction.
- Slide Graph is computationally more efficient than patch-based models and opens the avenue of using WSI graph representations for solving other problems in computational pathology as well.

2. Problem Formulation

Development of machine learning models that can capture tissue micro-architecture and geometry to generate WSI-level label predictions is an open problem in computational pathology. A typical machine learning problem in computational pathology involves a training dataset $\{(x_i, y_i) | i = 1 \dots M\}$ of M WSIs x_i each with a label $y_i \in \{0, +1\}$. The objective is then to develop a machine learning model such as a neural network based decision function $f(x_i; \theta)$, parameterized by its weight parameters θ , that can predict the label for unseen cases. In this paper, we consider two prediction problems: prediction of status of HER2 and PR from H&E stained images. HER2 is a growth-promoting biomarker/protein that helps breast cells grow, divide, and repair themselves and breast cancer cells that over-express HER2 are called HER2-positive. HER2-positive breast cancers tend to grow and spread faster than HER2-negative cancers, but are much more likely to respond to treatment with specific drugs [28, 29]. Similarly, PR is a valuable prognostic biomarker for determining survival, drug response and progression [30, 31]. Conventionally, immunohistochemistry (IHC) markers are used for determination of HER2 and PR status [32]. To the best of our knowledge, Kather et al. in [33] are the first to propose a deep learning method to predict hormone receptor status from routine H&E WSIs. However, like other existing approaches in computation pathology, their method is also based on patch-level prediction. As discussed below, we overcome these limitations using a graph-based method that can capture cellular organization and geometry in tissue slides – the so-called *histology landscape* at the entire WSI level.

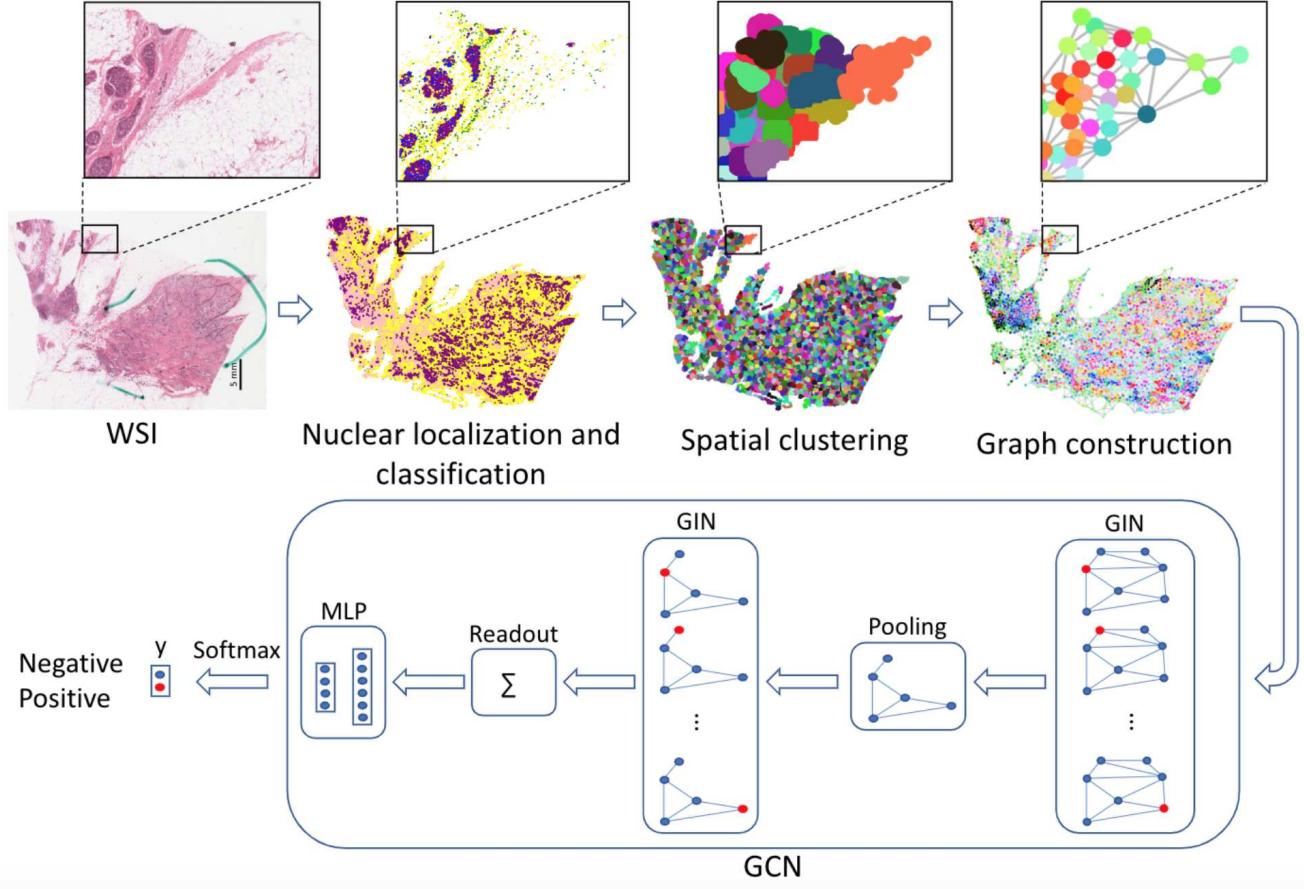


Figure 2. Workflow of proposed Slide Graph for graph classification. Four steps are needed: nuclear segmentation and classification; spatial clustering; graph construction and GCN.

3. The Proposed Method

The proposed Slide Graph method first builds a graph representation $G_i = G(x_i)$ of a WSI and then uses a graph convolutional neural network to generate slide level predictions $f(G(x_i; \theta))$. It consists of four steps (Fig. 2): first, we use HoVer-Net [34] to perform simultaneous nuclear segmentation and classification and extract nuclear features. Second, we use spatial clustering to group a set of spatially neighboring nuclei into clusters. Third, we develop a graph representation of the clusters to capture cellular topology of the WSI. Lastly, the graph built on the entire WSI is taken as an input to a graph convolutional neural network to predict receptor status at the WSI-level. Below, we present details of each of these steps.

3.1. Nuclear Segmentation and Classification

In order to construct a graph representation of the entire WSI, we first use HoVer-Net [34] to localize nuclei and predict their types. HoVer-Net is a convolutional neural network for simultaneous nuclear segmentation and classi-

fication. This network leverages instance-rich information encoded within vertical and horizontal distances of nuclear pixels to their centres of mass and achieves accurate segmentation even in areas with overlapping instances. We train HoVer-Net on the PanNuke dataset [35], which consists of 19 different tissue types, and then predict the following nuclei categories: neoplastic; non-neoplastic epithelial; inflammatory; connective tissue and dead. For a given WSI, this results in a set of N nuclei $P = \{p_j | j = 1 \dots N\}$ with nuclear centroids at p_j in conjunction with the type and morphological features of each nucleus. This set is then used for graph construction (see **Algorithm 1**).

3.2. Spatial clustering

A single WSI can contain hundreds of thousands of nuclei. In this paper, we use Agglomerative clustering [36] to group spatially neighboring nuclei into clusters to reduce the computational cost of downstream analysis. Specifically, we select a random subset $R \subseteq P$ of up to 10,000 nuclei for agglomerative clustering using Euclidean distance

metric with average linkage. This is done such that cluster agglomeration takes place up to a minimum distance threshold of $d_{min} = 500$ -pixel units. This results in a set of K clusters represented by the set $C = \{c_k | k = 1...K\}$ with each $\mathbf{p}_j \in R$ assigned to exactly one cluster. We then use the nearest neighbor rule [37] to assign all points in P to clusters in C , i.e., each point p_j is assigned the cluster membership of its nearest neighbor in R . This results in spatial clustering of all nuclei in a given WSI.

3.3. Graph construction

In this step, we construct a global planar graph representation $G = (V, E)$ [38, 39, 40] of a given WSI. For this purpose, we treat each cluster $c \in C$ as a node to construct a vertex set $V = \{v_c | c \in C\}$ such that each node is represented by the cluster's geometric center $\mathbf{g}_c = \frac{1}{|c|} \sum_{\mathbf{p}_j \in c} \mathbf{p}_j$ and a feature vector \mathbf{h}_c based on the type and morphology of its constituent nuclei. Specifically, in each cluster, we use the count of the six nuclei types and the standard deviation of nuclear sizes as features to capture local cellular heterogeneity. The edge set $E \subseteq V \times V$ represents a finite set of edges between nodes. In order to capture communication patterns between components of the tissue, the edge set is constructed by using Delauney triangulation based on the geometric coordinates of cluster centers with a maximum distance connectivity threshold of $d_{max} = 4,000$ pixels constrained. This results in a planar graph, i.e., no two edges in the graph intersect each other. The relationship between distant nodes can be modeled by the structure of the graph neural network itself.

3.4. WSI graph visualization

In order to understand the ability of WSI-level graphs to capture tissue architecture and their predictive power for WSI-level prediction of receptor status, we used Principal Component Analysis (PCA) [41] over node level features to visualize differences between HER2+ and HER2- WSIs. Specifically, PCA is first used to reduce node feature dimensionality from 7 to 3 and then a false color representation of each node is generated based on its PCA coefficients. This results in a WSI level graph visualization in which the color of each node is based on its node features whereas the location of the node represents its geometric center. Fig. 3 shows the results of this visualization for 5 HER2- (top row) and 5 HER2+ (bottom row) WSIs. Despite the use of unsupervised PCA, one can observe clear differences in the graphs of the two classes: note the prevalence of red areas in HER2+ WSIs and blue areas in HER2- WSIs. This supports the overall idea of using WSI level graphs for machine learning problems in computational pathology proposed in this work.

Algorithm 1: WSI Graph Construction.

INPUT: A set of N nuclei $P = \{\mathbf{p}_j | j = 1...N\}$ detected in a given WSI - each represented by its spatial coordinates \mathbf{p}_j .

OUTPUT: Graph representation $G = (V, E)$ of the WSI.

PARAMETERS:

$$n = 10,000, d_{min} = 500, d_{max} = 4,000$$

STEPS:

- 1: Select a random subset $R \subseteq P$ of up to n nuclei
 - 2: Perform agglomerative clustering based on R using Euclidean distance metric with average linkage. This is done such that cluster agglomeration takes place up to a minimum distance threshold d_{min} . This results in a set of K clusters represented by the set $C = \{c_k | k = 1...K\}$ with each $\mathbf{p}_j \in R$ assigned to exactly one cluster.
 - 3: Each point $\mathbf{p}_j \in P$ is assigned the cluster membership of its nearest neighbor in R .
 - 4: Compute the geometric center \mathbf{g}_c of each cluster set $c \in C$, i.e., without introducing further notation, $\mathbf{g}_c = \frac{1}{|c|} \sum_{\mathbf{p}_j \in c} \mathbf{p}_j$.
 - 5: Compute a feature representation \mathbf{h}_c of each cluster $c \in C$ based on type and morphology of its constituent nuclei.
 - 6: Construct a vertex set V for the given WSI consisting of nodes $v_c = (\mathbf{h}_c, \mathbf{g}_c) \in V$.
 - 7: Use Delauney triangulation to construct the edge set E based on geometric centers of clusters \mathbf{g}_c in the WSI with a maximum connectivity distance threshold of d_{max} .
-

3.5. Graph Isomorphic Neural Networks

After constructing WSI graphs, the task of WSI-level label prediction can be considered as a graph classification problem. We use a graph convolutional neural network (GCN) with graph isomorphic network convolutional (GIN-Conv) layers [42] for this purpose. The GCN can be used to generate predictions for an input graph representation G of a given WSI. The underlying structure of GINConv layers is based on the Weisfeiler-Lehman graph isomorphism test [43] which makes them particularly effective for graph classification [44], specially for problems in computational pathology as local cellular architecture in WSIs may be invariant to translation and rotational changes. The GCN used for WSI-level classification in the proposed method can be considered as a mechanism for learning progressively abstract representations of node level features based on their local neighborhood through message passing between neighboring nodes. Mathematically, the output feature representation of a GINConv layer l in the GCN for a given node k in the input graph can be represented by:

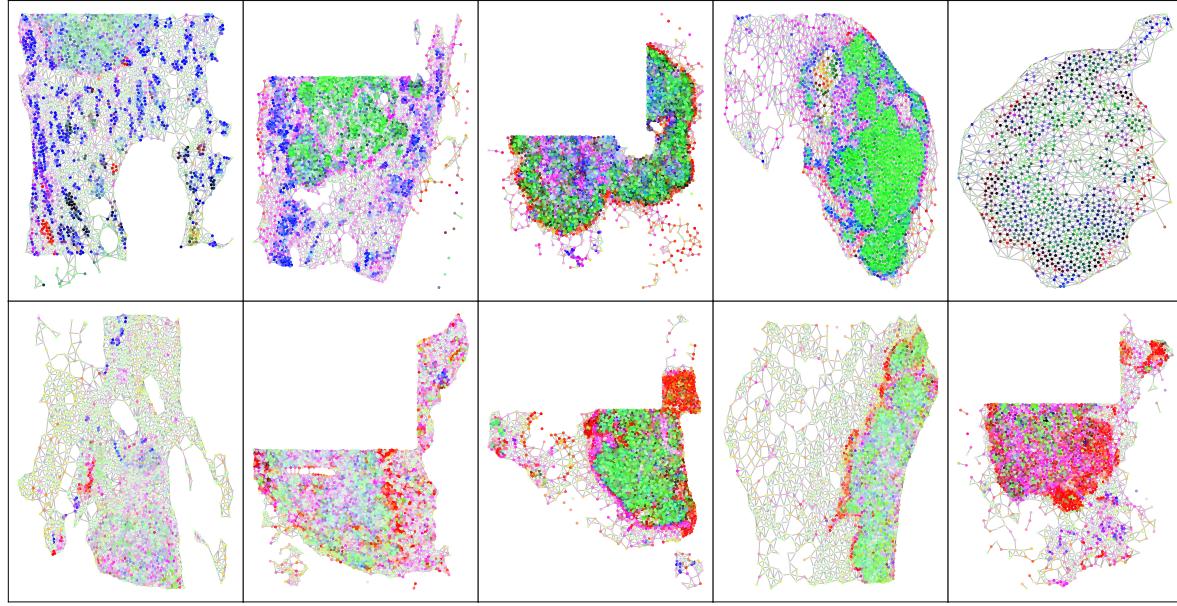


Figure 3. Visualization of nodes’ features. Top row: HER2 negative; Bottom row: HER2 positive.

$$\mathbf{h}_k^{(l)} = H_{\theta_l}^{(l)} \left(\left(1 + \varepsilon^{(l)} \right) \cdot \mathbf{h}_k^{(l-1)} + \sum_{u \in \mathcal{N}_k} \mathbf{h}_u^{(l-1)} \right), \quad (1)$$

Here, $\mathbf{h}_k^{(l)}$ is the feature vector of node k at the l -th convolutional layer, \mathcal{N}_k denotes the neighbourhood of node k , i.e., $\mathcal{N}_k = \{c \in V : (c, k) \in E\}$, and ε is a scalar parameter that determines the local impact of a node. $H_{\theta_l}^{(l)}$ represents a multi-layer perceptron (MLP) that learns a non-linear feature transformation at the node level for the l -th layer with weight parameters θ_l . At the input layer ($l = 0$), the network takes node level features discussed above as input, i.e., $\mathbf{h}_k^{(0)} = \mathbf{h}_k$. The GCN produces a graph-level prediction score $f(G(x))$ by using a readout layer that essentially accumulates node level feature representations through summation. The MLP weights in each layer are tuned through backpropagation of gradients over a weighted cross-entropy loss function [45] between model predictions and training labels. In this work, we use two GINConv layers ($l = 2$) which have 16 and 8 neurons respectively. The corresponding MLP is set as linear \rightarrow RELU \rightarrow linear \rightarrow BN (batch normalization) \rightarrow RELU. ε is a trainable parameter which is updated through the learning process.

3.6. Implementation and hyperparameter setting

The proposed Slide Graph is implemented ² using PyTorch Geometric (PyG) [46] [47]. For training, we use adaptive momentum based optimization (Adam [48]) with the learning rate of $5e^{-4}$ and a weight decay $1e^{-4}$.

²The code can be obtained via email request or something to that effect.

3.7. Performance Evaluation and Comparison

In order to compare with state-of-the-art methods [33], we evaluate the performance of our proposed network on the same H&E stained cohort from The Cancer Genome Atlas in breast cancer (TCGA-BRCA) [49]. This dataset totally has 709 WSIs. Among them, in HER2 status differentiation, there are 608 HER2 negative and 101 HER2 positive images while in PR status differentiation, 452 PR positive and 256 PR negative images are included. In order to deal with the imbalance between the classes in the datasets, we use a weighted cross-entropy loss function where the minority class is assigned a large weight in the loss function. We use the area under the receiver operating characteristic curve (AUC-ROC) as a performance metric. We used three stratified fold cross-validation for a direct comparison with two patch-based classification methods which are explained below [33, 50].

3.7.1 Method 1

To the best of our knowledge, Kather *et al.* [33] is the first and only publication to propose a deep learning method to predict HER2 status from H&E stained tissue images. This work uses a pretrained ShuffleNet to learn the patch based features. During the training and testing process, only 1000 patches are randomly selected from each WSI. The patient cohort is randomly split in three parts in the way that each part contained approximately the same number of patients with each label. These three parts of the patient cohort were then used for three-fold patient-level cross-validation. The

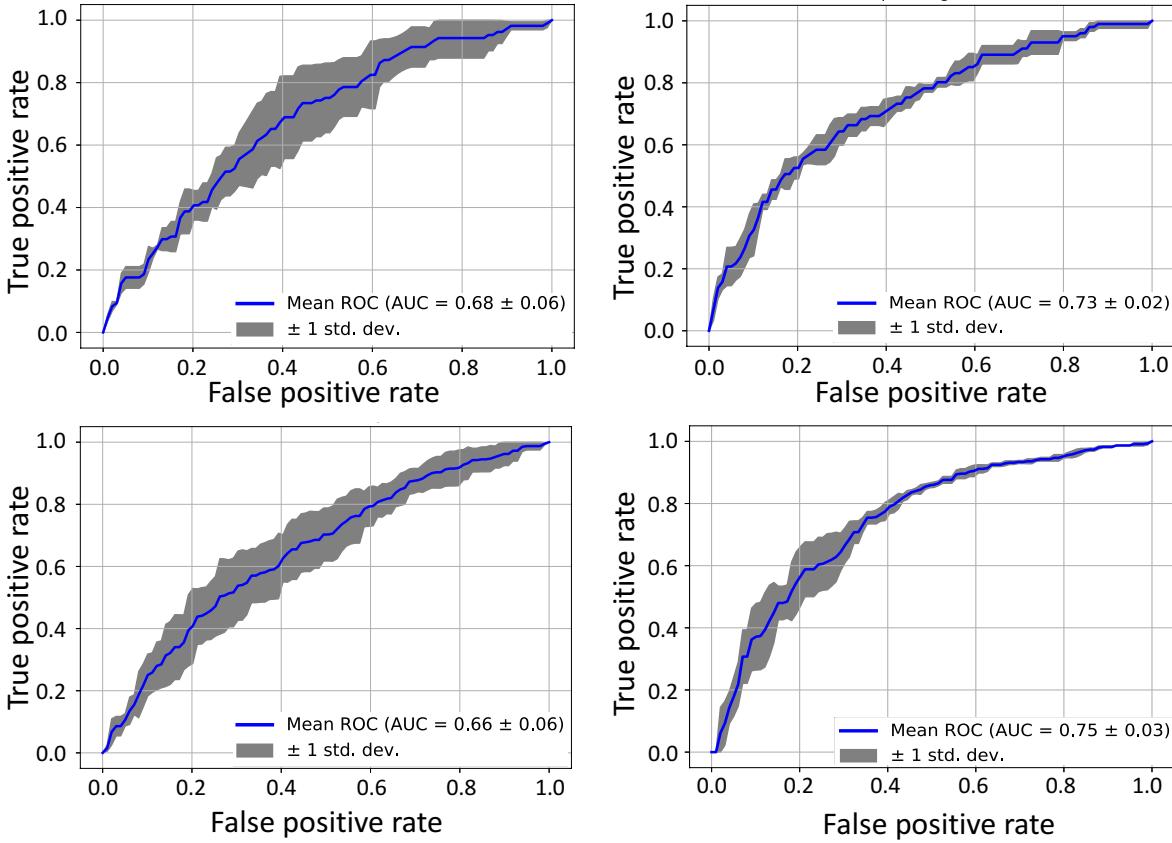


Figure 4. AUC-ROC using different methods (Left column: deep learning based model proposed in [50]; right: our proposed Slide Graph) for HER2 (top row) and PR (bottom row) status prediction.

area under the receiver operating characteristic curve (AUC-ROC) achieved is 0.62 for HER2 and 0.73 for PR status prediction [33].

3.7.2 Method 2

This work is our own implementation of the deep learning based model originally published by Kather *et al.* [50] for the prediction of microsatellite instability in colorectal cancer patients. A pretrained ResNet18 has been used to learn image based features from the non-overlapping square tiles of each whole slide image. The trained model gives a score to each tile and then a majority voting based aggregation method is finally used to obtain a whole slide prediction. We use this model as a baseline method to predict receptor status for model comparison.

4. Results and Discussion

The results of three fold cross-validation over both prediction problems are given in Table 1 and Fig. 4. The currently published state of the art method by Kather *et al.* [33] gives three-fold cross-validation mean AUC-ROC scores of

0.62 and 0.73, for HER2 and PR status prediction, respectively³. The proposed Slide Graph model gives better predictive performance in comparison to the existing state of the art by a significant margin: 0.62 to 0.73 in HER2 and 0.73 to 0.75 in PR status prediction. In addition, comparing with the method in [50], our proposed model succeeds with higher AUC-ROC score and smaller standard deviation, which illustrates the stability of our proposed method. The confusion matrices for the two classification problems are given in Table 2.

Table 1. AUC-ROC using different models on HER2/PR status prediction (Mean ± Standard deviation).

AUC-ROC	HER2	PR
Method 1 [33]	0.62	0.73
Method 2 [50]	0.68 ± 0.06	0.66 ± 0.06
Slide Graph	0.73 ± 0.02	0.75 ± 0.03

We have also compared the computational efficiency of patch-based and the proposed Slide Graph models using a single NVIDIA TITAN RTX GPU. Once the patches and

³ROC curves and standard deviation values are not reported in [33].

Table 2. Confusion matrix for receptor status prediction. Top: HER2; bottom: PR.

n=231	Predicted negative	Predicted positive
True negative	120	77
True positive	7	27
n=236	Predicted negative	Predicted positive
True negative	51	34
True positive	31	120

graphs ready from the WSI, the average single fold training time for the baseline model [50] is 5.3 hours and the testing time for a WSI is 1.2 seconds from patches to the final prediction whereas Slide Graph training for a single fold takes 3 minutes on average and 0.4 milliseconds to get the label prediction from a single graph. These results clearly show that Slide Graph is well suited to predict the receptor status from H&E stained images by capturing geometric cellular structure in WSIs.

5. Conclusions

In this paper, we proposed a novel WSI level representation, termed **Slide Graph**, that can be easily coupled with the graph convolutional networks, for label prediction directly from the WSI. This method can effectively overcome the drawbacks of patch-based methods by capturing the biological geometric structure of the cellular architecture at the entire WSI level. The proposed Slide Graph can effectively incorporate both cell-level and contextual information by using nuclear features and graph convolution. Experimental analysis over clinically important tasks of HER2 and PR receptor status prediction show that the proposed Slide Graph method can produce higher accuracy than previous state-of-the-art techniques. Slide Graph can be applied to other problems in computational pathology, such as recurrence and survival prediction.

Acknowledgements

WL, SG, NR and FM are supported by the PathLAKE digital pathology consortium, which is funded from the Data to Early Diagnosis and Precision Medicine strand of the government’s Industrial Strategy Challenge Fund, managed and delivered by UK Research and Innovation (UKRI). MB is funded by the UK Medical Research Council (No. MR/P015476/1).

References

- [1] Thomas J Fuchs, Peter J Wild, Holger Moch, and Joachim M Buhmann. Computational pathology analysis of tissue microarrays predicts survival of renal clear cell carcinoma patients. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 1–8. Springer, 2008. 1
- [2] Anant Madabhushi and George Lee. Image analysis and machine learning in digital pathology: Challenges and opportunities, 2016. 1
- [3] Farzad Ghaznavi, Andrew Evans, Anant Madabhushi, and Michael Feldman. Digital imaging in pathology: whole-slide imaging and beyond. *Annual Review of Pathology: Mechanisms of Disease*, 8:331–359, 2013. 1
- [4] Alexander JJ Smits, J Alain Kummer, Peter C De Bruin, Mijke Bol, Jan G Van Den Tweel, Kees A Seldenrijk, Stefan M Willems, G Johan A Offerhaus, Roel A De Weger, Paul J Van Diest, et al. The estimation of tumor cell percentage for molecular testing by pathologists is not accurate. *Modern Pathology*, 27(2):168–174, 2014. 1
- [5] Hollis Viray, Kevin Li, Thomas A Long, Patricia Vasalos, Julia A Bridge, Lawrence J Jennings, Kevin C Halling, Meera Hameed, and David L Rimm. A prospective, multi-institutional diagnostic trial to determine pathologist accuracy in estimation of percentage of malignant cells. *Archives of Pathology and Laboratory Medicine*, 137(11):1545–1549, 2013. 1
- [6] Neeraj Kumar, Ruchika Verma, Sanuj Sharma, Surabhi Bhargava, Abhishek Vahadane, and Amit Sethi. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE transactions on medical imaging*, 36(7):1550–1560, 2017. 1
- [7] Le Hou, Dimitris Samaras, Tahsin M Kurc, Yi Gao, James E Davis, and Joel H Saltz. Patch-based convolutional neural network for whole slide tissue image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2424–2433, 2016. 1, 2
- [8] Hamid Reza Tizhoosh and Liron Pantanowitz. Artificial intelligence and digital pathology: challenges and opportunities. *Journal of pathology informatics*, 9, 2018. 1
- [9] Andrew Janowczyk and Anant Madabhushi. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of pathology informatics*, 7, 2016. 1
- [10] Peter Bandi, Oscar Geessink, Quirine Manson, Marcory Van Dijk, Maschenka Balkenhol, Meyke Hermsen, Babak Ehteshami Bejnordi, Byungjae Lee,

- Kyunghyun Paeng, Aoxiao Zhong, et al. From detection of individual metastases to classification of lymph node status at the patient level: the camelyon17 challenge. *IEEE transactions on medical imaging*, 38(2):550–560, 2018. 1
- [11] Angel Cruz-Roa, Ajay Basavanhally, Fabio González, Hannah Gilmore, Michael Feldman, Shridar Ganesan, Natalie Shih, John Tomaszewski, and Anant Madabhushi. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In *Medical Imaging 2014: Digital Pathology*, volume 9041, page 904103. International Society for Optics and Photonics, 2014. 2
- [12] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. 2
- [13] Minh Hoai Nguyen, Lorenzo Torresani, Fernando De La Torre, and Carsten Rother. Weakly supervised discriminative localization and classification: a joint learning process. In *2009 IEEE 12th International Conference on Computer Vision*, pages 1925–1932. IEEE, 2009. 2
- [14] Stuart Andrews, Thomas Hofmann, and Ioannis Tsochantaridis. Multiple instance learning with generalized support vector machines. In *AAAI/IAAI*, pages 943–944, 2002. 2
- [15] Cha Zhang, John C Platt, and Paul A Viola. Multiple instance boosting for object detection. In *Advances in neural information processing systems*, pages 1417–1424, 2006. 2
- [16] Qi Zhang and Sally A Goldman. Em-dd: An improved multiple-instance learning technique. In *Advances in neural information processing systems*, pages 1073–1080, 2002. 2
- [17] Gabriele Campanella, Matthew G Hanna, Luke Geneslaw, Allen Mirafiori, Vitor Werneck Krauss Silva, Klaus J Busam, Edi Brogi, Victor E Reuter, David S Klimstra, and Thomas J Fuchs. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature medicine*, 25(8):1301–1309, 2019. 2
- [18] Judith MS Prewitt. Graphs and grammars for histology: An introduction. In *Proceedings of the Annual Symposium on Computer Application in Medical Care*, page 18. American Medical Informatics Association, 1979. 2
- [19] Cigdem Gunduz, Bülent Yener, and S Humayun Gültakin. The cell graphs of cancer. *Bioinformatics*, 20(suppl_1):i145–i151, 2004. 2
- [20] Cigdem Demir, S Humayun Gültakin, and Bülent Yener. Augmented cell-graphs for automated cancer diagnosis. *Bioinformatics*, 21(suppl_2):ii7–ii12, 2005. 2
- [21] Barbara Weyn, Gert van de Wouwer, Samir Kumar Singh, André van Daele, Paul Scheunders, Eric Van Marck, and Willem Jacob. Computer-assisted differential diagnosis of malignant mesothelioma based on syntactic structure analysis. *Cytometry: The Journal of the International Society for Analytical Cytology*, 35(1):23–29, 1999. 2
- [22] L Paul Chew. Constrained delaunay triangulations. *Algorithmica*, 4(1-4):97–108, 1989. 2
- [23] Sahirzeeshan Ali, Robert Veltri, Jonathan A Epstein, Christhunesa Christudass, and Anant Madabhushi. Cell cluster graph for prediction of biochemical recurrence in prostate cancer patients from tissue microarrays. In *Medical Imaging 2013: Digital Pathology*, volume 8676, page 86760H. International Society for Optics and Photonics, 2013. 2
- [24] Bülent Yener. Cell-graphs: image-driven modeling of structure-function relationship. *Communications of the ACM*, 60(1):74–84, 2016. 2
- [25] Cheng Lu, Xiangxue Wang, Prateek Prasanna, German Corredor, Geoffrey Sedor, Kaustav Bera, Vamsidhar Velcheti, and Anant Madabhushi. Feature driven local cell graph (fedeg): Predicting overall survival in early stage lung cancer. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 407–416. Springer, 2018. 2
- [26] Jon Whitney, German Corredor, Andrew Janowczyk, Shridar Ganesan, Scott Doyle, John Tomaszewski, Michael Feldman, Hannah Gilmore, and Anant Madabhushi. Quantitative nuclear histomorphometry predicts oncotype dx risk categories for early stage er+ breast cancer. *BMC cancer*, 18(1):610, 2018. 2
- [27] Korsuk Sirinukunwattana, David Snead, David Epstein, Zia Aftab, Imaad Mujeeb, Yee Wah Tsang, Ian Cree, and Nasir Rajpoot. Novel digital signatures of tissue phenotypes for predicting distant metastasis in colorectal cancer. *Scientific reports*, 8(1):1–13, 2018. 2
- [28] Yosef Yarden. Biology of her2 and its importance in breast cancer. *Oncology*, 61(Suppl. 2):1–13, 2001. 2

- [29] Rita Nahta, Dihua Yu, Mien-Chie Hung, Gabriel N Hortobagyi, and Francisco J Esteva. Mechanisms of disease: understanding resistance to her2-targeted therapy in human breast cancer. *Nature clinical practice Oncology*, 3(5):269–280, 2006. 2
- [30] Lisa K Dunnwald, Mary Anne Rossing, and Christopher I Li. Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. *Breast cancer research*, 9(1):R6, 2007. 2
- [31] Xiaojiang Cui, Rachel Schiff, Grazia Arpino, C Kent Osborne, and Adrian V Lee. Biology of progesterone receptor loss in breast cancer and its implications for endocrine therapy. *Journal of clinical oncology*, 23(30):7721–7735, 2005. 2
- [32] Talha Qaiser, Abhik Mukherjee, Chaitanya Reddy Pb, Sai D Munugoti, Vamsi Tallam, Tomi Pitkähö, Taina Lehtimäki, Thomas Naughton, Matt Berseth, Aníbal Pedraza, et al. Her 2 challenge contest: a detailed assessment of automated her 2 scoring algorithms in whole slide images of breast cancer tissues. *Histopathology*, 72(2):227–238, 2018. 2
- [33] Jakob Nikolas Kather, Lara R Heij, Heike I Grabsch, Loes FS Kooreman, Chiara Loeffler, Amelie Echle, Jeremias Krause, Hannah Sophie Muti, Jan M Niehues, Kai AJ Sommer, et al. Pan-cancer image-based detection of clinically actionable genetic alterations. *bioRxiv*, page 833756, 2019. 2, 5, 6
- [34] Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, 58:101563, 2019. 3
- [35] Jevgenij Gamper, Navid Alemi Koohbanani, Ksenija Benet, Ali Khuram, and Nasir Rajpoot. Pannuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification. In *European Congress on Digital Pathology*, pages 11–19. Springer, 2019. 3
- [36] Daniel Müllner. Modern hierarchical, agglomerative clustering algorithms. *arXiv preprint arXiv:1109.2378*, 2011. 3
- [37] James M Keller, Michael R Gray, and James A Givens. A fuzzy k-nearest neighbor algorithm. *IEEE transactions on systems, man, and cybernetics*, (4):580–585, 1985. 4
- [38] Harshita Sharma, Norman Zerbe, Sebastian Lohmann, Klaus Kayser, Olaf Hellwich, and Peter Hufnagl. A review of graph-based methods for image analysis in digital histopathology. *Diagnostic pathology*, 1(1), 2015. 4
- [39] Wenqi Lu, Jinming Duan, Joshua Deepak Veesa, and Iain B Styles. New nonlocal forward model for diffuse optical tomography. *Biomedical Optics Express*, 10(12):6227–6241, 2019. 4
- [40] Wenqi Lu, Jinming Duan, David Orive-Miguel, Lionel Herve, and Iain B Styles. Graph-and finite element-based total variation models for the inverse problem in diffuse optical tomography. *Biomedical optics express*, 10(6):2684–2707, 2019. 4
- [41] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987. 4
- [42] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018. 4
- [43] Boris Weisfeiler and Andrei A Lehman. A reduction of a graph to a canonical form and an algebra arising during this reduction. *Nauchno-Technicheskaya Informatsiya*, 2(9):12–16, 1968. 4
- [44] László Babai and Ludik Kucera. Canonical labelling of graphs in linear average time. In *20th Annual Symposium on Foundations of Computer Science (sfcs 1979)*, pages 39–46. IEEE, 1979. 4
- [45] Zhilu Zhang and Mert Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. In *Advances in neural information processing systems*, pages 8778–8788, 2018. 5
- [46] Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*, 2019. 5
- [47] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017. 5
- [48] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5
- [49] Cancer Genome Atlas Network et al. Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418):61, 2012. 5
- [50] Jakob Nikolas Kather, Alexander T Pearson, Niels Haima, Dirk Jäger, Jeremias Krause, Sven H Loosen,

Alexander Marx, Peter Boor, Frank Tacke, Ulf Peter Neumann, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nature medicine*, 25(7):1054–1056, 2019.

[5](#), [6](#), [7](#)