Medical Image Analysis 80 (2022) 102486

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

SlideGraph+: Whole slide image level graphs to predict HER2 status in breast cancer

Wenqi Lu[a], Michael Toss[b], Muhammad Dawood[a], Emad Rakha[b], Nasir Rajpoot[a], Fayyaz Minhas[a,*]

[a] Tissue Image Analytics (TIA) Centre, Department of Computer Science, University of Warwick, UK
[b] Nottingham Breast Cancer Research Centre, Division of Cancer and Stem Cells, School of Medicine, Nottingham City Hospital, University of Nottingham, Nottingham, UK

ARTICLE INFO

Article history:
Received 8 October 2021
Revised 15 April 2022
Accepted 20 May 2022
Available online 25 May 2022

Keywords:
Computational pathology
Breast cancer
Weak supervision
Human epidermal growth factor receptor
Graph neural networks

ABSTRACT

Human epidermal growth factor receptor 2 (HER2) is an important prognostic and predictive factor which is overexpressed in 15–20% of breast cancer (BCa). The determination of its status is a key clinical decision making step for selection of treatment regimen and prognostication. HER2 status is evaluated using transcriptomics or immunohistochemistry (IHC) through in-situ hybridisation (ISH) which incurs additional costs and tissue burden and is prone to analytical variabilities in terms of manual observational biases in scoring. In this study, we propose a novel graph neural network (GNN) based model (SlideGraph+) to predict HER2 status directly from whole-slide images of routine Haematoxylin and Eosin (H&E) stained slides. The network was trained and tested on slides from The Cancer Genome Atlas (TCGA) in addition to two independent test datasets. We demonstrate that the proposed model outperforms the state-of-the-art methods with area under the ROC curve (AUC) values > 0.75 on TCGA and 0.80 on independent test sets. Our experiments show that the proposed approach can be utilised for case triaging as well as pre-ordering diagnostic tests in a diagnostic setting. It can also be used for other weakly supervised prediction problems in computational pathology. The SlideGraph+ code repository is available at https://github.com/wenqi006/SlideGraph along with an IPython notebook showing an end-to-end use case at https://github.com/TissueImageAnalytics/tiatoolbox/blob/develop/examples/full-pipelines/slide-graph.ipynb.

1. Introduction

Breast cancer (BCa) is the most commonly diagnosed cancer among women, and is the second leading cause of female cancer related deaths worldwide (Ahmad, 2019). Human epidermal growth factor receptor 2 (HER2) positivity accounts for around 15% of the early stage BCa. HER2 positivity in BCa is defined as evidence of HER2 protein overexpression and/or HER2 gene amplification (Ross et al., 2009) which is proven to be associated with worse clinical outcome (Slamon et al., 1987). HER2-positive BCa tumours tend to grow and spread faster than HER2-negative tumours, but are much more likely to respond to targeted therapy with anti-HER2 drugs (Yarden, 2001; Nahta et al., 2006).

In routine diagnostic practice, BCa tissue sections are stained with Haematoxylin and Eosin (H&E) and visually examined for morphological assessment. It is then followed by ancillary techniques including immunohistochemistry (IHC) and in situ hybridisation (ISH) to assess the expression of specific proteins, including HER2, for prognostic and predictive purposes (Fig. 1(a)). The current guidelines (Wolff et al., 2018) revised by the American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) assign a HER2 positivity score between 0 and 3+ based on visual analysis of IHC slides. Cases scoring 0 or 1+ are classified as HER2-negative (HER2-), while cases with a score of 3+ are regarded as HER2-positive (HER2+). Cases with score 2+ refer to equivocal expression of HER2 that need further assessment using ISH to evaluate HER2 gene status. Operational and analytical limitations of aforementioned techniques in terms of cost, tissue usability and observer-subjectivity in manual scoring affect interpretation of HER2 status and hence patient management. Consequently,

* Corresponding author.
E-mail address: Fayyaz.Minhas@warwick.ac.uk (F. Minhas).

https://doi.org/10.1016/j.media.2022.102486

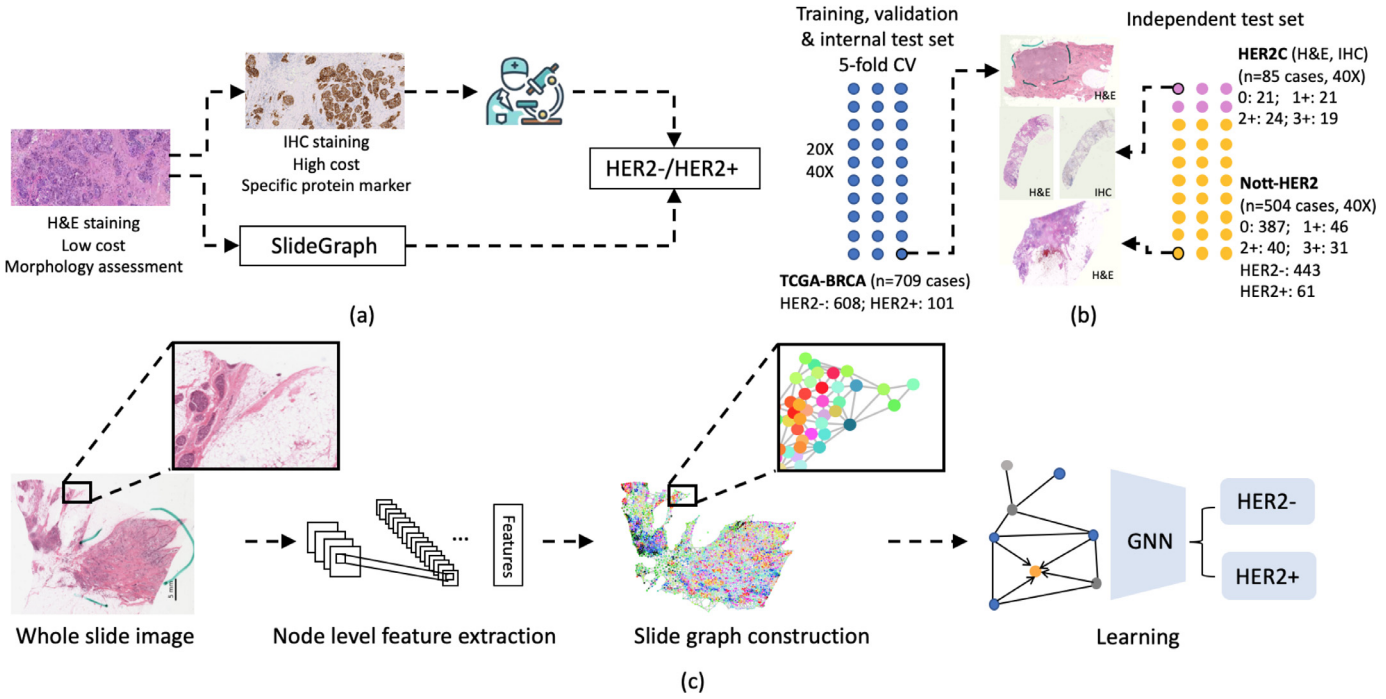

Fig. 1. HER2 status prediction from H&E images: (a) In routine diagnostic practice of BCa, tissue sections are commonly stained with H&E, followed by immunohistochemistry (IHC) staining to estimate the presence of specific protein receptors. We show that our deep learning algorithm can predict HER2 directly from H&E images; (b) Multi-centre datasets that are used to train, validate and test our proposed model; (c) Without pixel-level annotations, our algorithm is trained on the WSI-level graphs and predicts HER2 status directly.

prediction of HER2 status directly from digitally scanned whole slide images (WSIs) of routine H&E-stained tissue sections through deep learning or Artificial Intelligence (AI) techniques is of significant clinical and scientific interest.

Digital pathology and AI offer significant potential to overcome the aforementioned limitations and improve diagnostic consistency (Acs et al., 2020; Farahmand et al., 2022; Qaiser et al., 2018). Such computational pathology (CPath) models have been used for diagnostics as well as prediction of genetic expression correlates. Kather et al. (2019a) proposed a deep learning method to predict hormone receptor status from routine H&E WSIs. Morphological correlates of specific mutations have also been observed in H&E stained BCa histology images. Rawat et al. (2020) introduced the concept of "tissue fingerprints" to learn H&E features that can distinguish one patient from another. However, a major limitation of existing AI methods stems from patch-level analysis employed by these methods. As an entire WSI at full-resolution can be of the order of 150,000 × 100,000 pixels, training a model on the full-resolution WSIs is computationally challenging and expensive. A two-step patch-level approach is typically used to deal with large size WSIs (Fig. 2) (Janowczyk and Madabhushi, 2016; Bandi et al., 2018). First, the image is divided into small image tiles (or patches), where each patch is processed independently by the neural network (Tizhoosh and Pantanowitz, 2018). Then predicted scores for each patch within the WSI are aggregated into a WSI-level score, usually by pooling their results with various aggregating strategies such as average pooling, max pooling and majority voting (Cruz-Roa et al., 2014; LeCun et al., 1998; Nguyen et al., 2009).

Patch-level analysis used in conventional CPath models has two major drawbacks. First, local patches have limited visual context. The optimal resolution and patch size for analysis are highly problem-dependent (Hou et al., 2016). Image patches at a high magnification level lead to loss of contextual information whereas patches at lower magnification levels may not capture cell-level

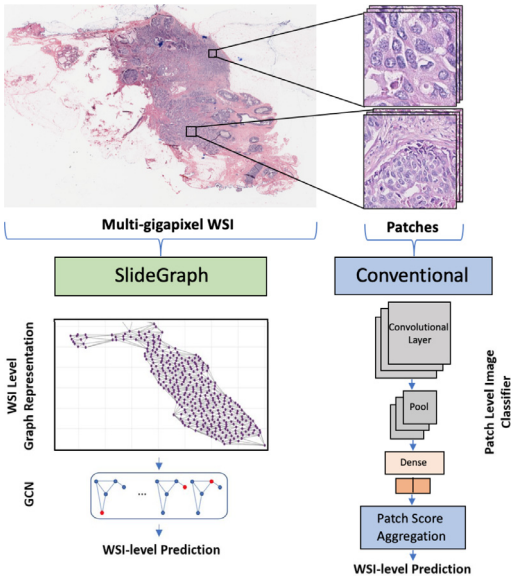

Fig. 2. The proposed SlideGraph model vs. conventional patch-based methods. Our proposed SlideGraph model is able to capture the overall organisation and structure of the tissue.

features (Fig. 2). Consequently, a patch-level machine learning method cannot capture the overall organisation and structure of the tissue in a WSI. Second, in most prediction problems in computational pathology, only WSI-level labels are available. It is non-trivial to model the association of each patch with a specific label. Weakly supervised machine learning methods such as Multiple Instance Learning (MIL) have been proposed to alleviate these problems about labels of training patches and aggregate patch-level predictions into WSI-level classification (Andrews et al., 2002;

Zhang et al., 2006; Zhang and Goldman, 2002; Campanella et al., 2019; Kather et al., 2019b; Lu et al., 2021; Li et al., 2021). However, these methods are unable to model the overall organisation and structure of the tissue at both global and local levels. As a consequence, graph-based approaches in computational pathology present a more principled way of modelling such prediction tasks.

The cell-graph technique (Prewitt, 1979; Gunduz et al., 2004; Demir et al., 2005; Jaume et al., 2021) was introduced to learn the structure-function relationship by modelling geometric structure of the tissue using graph theory. It is based on the assumption that cells in a tissue can organise in a certain way for specific functional states. Such cell-graphs can have different types, such as Delaunay triangles (Weyn et al., 1999; Chew, 1989), Voronoi diagrams, Minimum Spanning Trees (MST), and Cell Cluster Graphs (CCG) (Ali et al., 2013). Yener (2016) 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 characterise spatial proximity of histopathological primitives in tasks, such as survival prediction in lung cancer (Lu et al., 2018), risk stratification in BCa (Whitney et al., 2018) and distant metastasis prediction in colorectal cancer (Sirinukunwattana et al., 2018). However, all these graph-based methods with deep learning classifiers were trained on image patches which have limited visual context. In addition, since these methods have not been applied on the WSI level, extra patch-based voting methods are necessary to predict the label of a given WSI.

In this study, we propose a graph neural network model, termed as SlideGraph$^+$ to address the 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 novel pipeline which operates on a graph at the entire WSI-level for prediction of HER2 status. Specifically, we model the patch-level features as a graph that can capture both cell-level and contextual information and does not require any patch-level labels. A graph neural network is then used for WSI-level prediction. This work is a significant extension of our previously proposed SlideGraph (Lu et al., 2020) and employs an extended network architecture that makes the model more interpretable. Outputs from the proposed network not only cover the overall prediction score, but also show the active graph nodes corresponding to image regions which contribute to the overall prediction. The proposed SlideGraph$^+$ model also incorporates a novel message passing technique and a modified loss function. 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 clinically relevant prediction problems from BCa H&E WSIs. Specifically, we train a classification model to predict the status of HER2 and test it on another two independent cohorts (Fig. 1(b)). Overall, our main contributions in this paper can be summarised as follows:

- SlideGraph is the first method which can generate whole slide image level predictions by using a graph representation of the cellular interconnection geometry in a WSI.
- The proposed SlideGraph$^+$ network architecture is an extension of our previously proposed SlideGraph (Lu et al., 2020) with an architecture layout which makes the network more interpretable by generating node-level predictions.
- SlideGraph$^+$ makes use of nuclear composition, nuclear morphology, neural network embeddings or DAB density estimates features to represent the complex organisation of cells and the overall tissue micro-architecture. The proposed network outperforms the state-of-the-art methods by a significant margin in HER2 status prediction.
- The DAB density regression model proposed in this paper is the first method to predict DAB intensity directly from H&E stained images. It carries potential of removing the necessity of IHC staining when evaluating the HER2 expression.
- Instead of annotating invasive tumour regions which is very time-consuming, SlideGraph$^+$ is trained on all tissue regions and is able to precisely localise the regions that contribute to the HER2 positivity and expression.
- Our trained HER2 status prediction model is tested on two independent cohorts, demonstrating its generalisation on multi-centre datasets.
- SlideGraph is computationally more efficient than patch-based models and opens the avenue of using WSI graph representations for solving other problems in computational pathology.

2. Methodology

The proposed framework for predicting the receptor status from H&E images is shown in Fig. 1(c). A typical weakly supervised machine learning problem in computational pathology involves a training dataset $\{(X_i, y_i)|i = 1 \dots M\}$ of M WSIs denoted by X_i, each with a label $y_i \in \{0, 1\}$. The objective is then to develop a machine learning model that can predict the label for unseen cases. In this work, we build a graph representation $G_i = G(X_i)$ of each X_i in the training set and train a graph neural network with trainable parameters θ to generate slide-level predictions $F(G(X_i); \theta)$. The trained model F is used for inference to predict status for WSIs which are not included in the training set.

The overall framework consists of four steps: first, we extract features from local regions in the WSI after preprocessing. Specifically, a given WSI X is modelled as a set of image patches $x_j \in X$ of size 512 $\times$ 512 pixels at 40$\times$ magnification. Each patch $x_j \equiv \left(\boldsymbol{g}_j, \boldsymbol{h}_j\right)$ is represented as a tuple consisting of a d-dimensional feature vector representation $\boldsymbol{h}_j \in \mathbb{R}^d$ and the corresponding geometric coordinates $\boldsymbol{g}_j \in \mathbb{R}^2$ of the top-left corner of the patch. Second, we use spatial clustering to group neighbouring image patches with similar features into clusters. Third, a graph representation based on these clusters is generated to capture the cellular and morphological topology of the WSI. Finally, the graph constructed from the entire WSI is taken as an input to a graph neural network to predict the receptor status at the graph node-level and also at the slide-level. Below, we give details of the datasets and individual steps in the proposed pipeline.

2.1. Datasets

The training dataset used in this study was obtained from The Cancer Genome Atlas in breast cancer (TCGA-BRCA) (Network et al., 2012). Molecular status of HER2 was assessed clinically on the patient level. We used five-fold stratified cross-validation for a direct comparison with other patch-based classification methods (Rawat et al., 2020; Kather et al., 2019a). In each fold-run, 20% of the dataset (at the patient level) was held out as unseen test data whereas the remaining 80% was used for training and validation. We then tested the trained model on other two independent cohorts, the publicly available HER2Contest challenge (HRE2C) dataset (Qaiser et al., 2018) and an internal Nottingham University Hospital (Nott-HER2) dataset, respectively. Some high-level information of all three datasets is shown in Fig. 1(b).

2.2. Pre-processing

We use stain normalisation technique by Vahadane et al. (2016) to normalise the stain distribution across slides especially those from different centres. Tissue segmentation

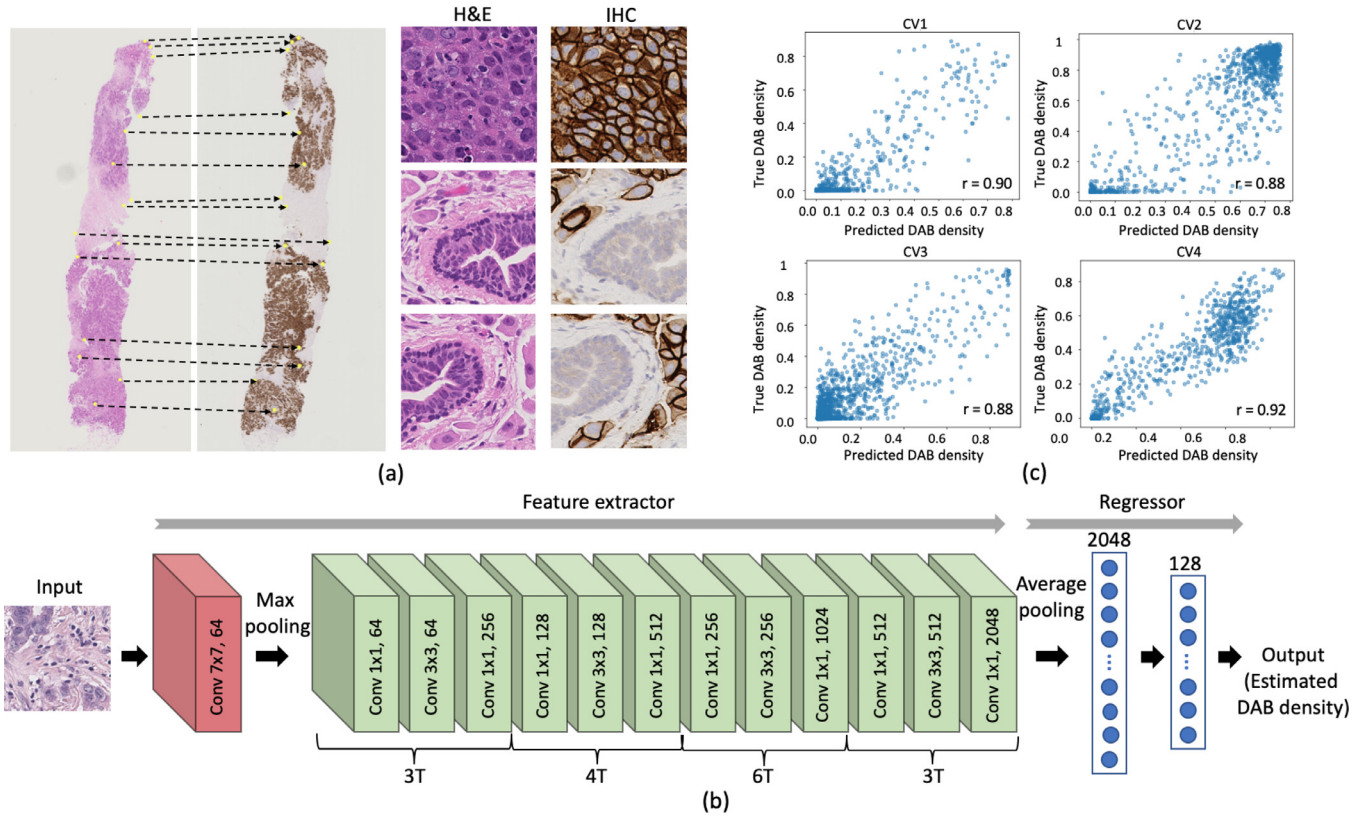

Fig. 3. DAB density estimation using a small subset of the HER2C dataset (Qaiser et al., 2018): (a) Image registration between H&E and corresponding IHC images using control points (yellow points *left*) on both WSIs; *right*: three examples of H&E patches and corresponding registered IHC patches; (b) Convolutional neural network architecture for regressing DAB density from H&E images. 'T' represents the number of resnet blocks used; (c) Scatter plots between model prediction and true DAB density using 4-fold cross validation (averaged Pearson correlation coefficient 0.90 with p-value < 0.0001).

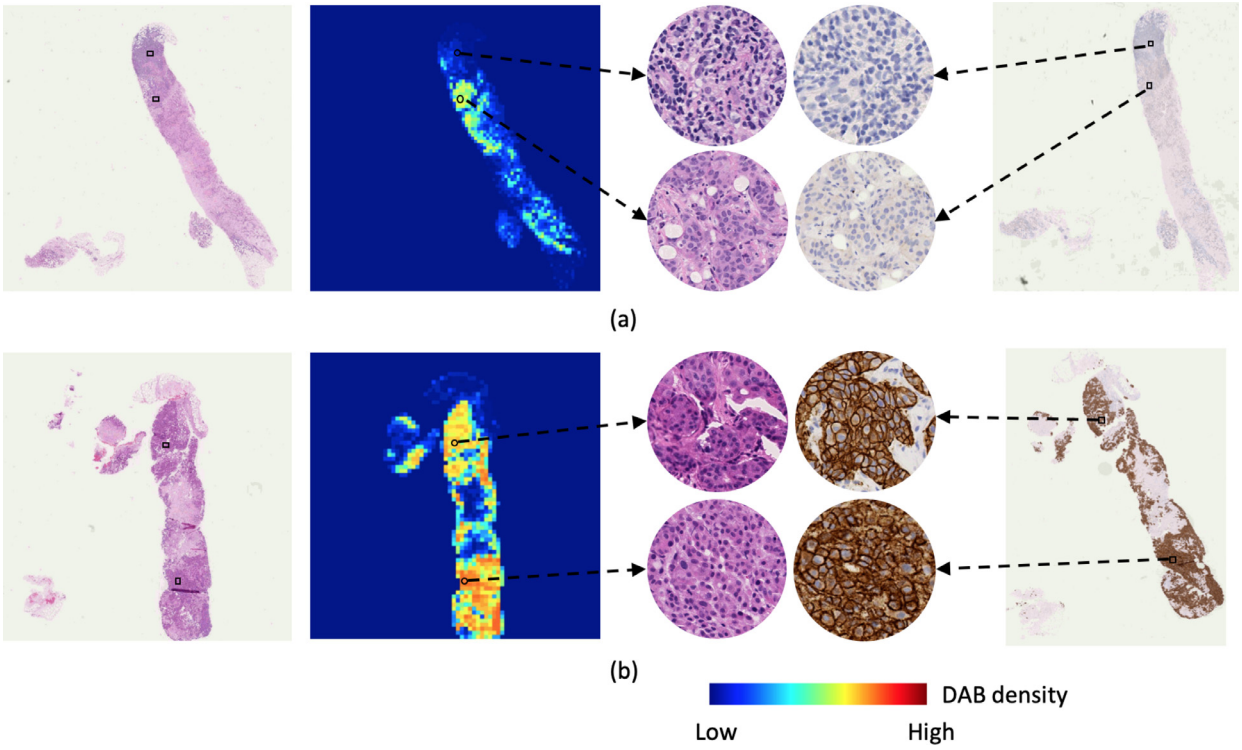

Fig. 4. Visualisation of the estimated DAB density on test WSIs from the HER2C dataset (Qaiser et al., 2018): (a) HER2-negative case; (b) HER2-positive case. From left to right column: raw H&E stained WSI; estimated DAB density using the trained regression model; zoomed-in version of two local regions; corresponding IHC stained WSI as the reference.

is performed to remove background regions. The segmented tissue region is then divided into a set of uniformly sized patches.

2.3. Feature extraction

Given a stain normalised patch x_j, a variety of representative features $\boldsymbol{h}_j$ can be extracted. The objective of this feature extraction step is to obtain features that are associated with tissue characteristics in the patch and the target variable of interest as well. These features include nuclear composition features (e.g., counts of different types of nuclei in the patch), morphological features, receptor expression features (Section 2.3.4), deep features (or neural feature embdeddings from a pre-trained neural network) and a combination of these. These features are then used for construction of the WSI-level graph. We explore the association between different kinds of features and the HER2 status. It is important to note that the proposed framework is generic and not restricted to any particular type of features and the graph neural network (discussed later) can be used to accumulate a variety of node level features.

2.3.1. Nuclear composition features (NCF)

HER2 status has been shown to be associated with the presence of different types of nuclei in BCa tissues (Lu et al., 2020). Here, we extract nuclear composition features which cover the counts of nuclei of different types of cells in a patch. Specifically, for a given patch x_j, we use HoVer-Net (Graham et al., 2019) trained on the BCa PanNuke dataset (Gamper et al., 2019) to localise nuclei and predict their types. HoVer-Net is a convolutional neural network for simultaneous nuclear segmentation and classification. This network leverages instance-rich information encoded within vertical and horizontal distance maps of nuclear pixels to their centres of mass and achieves accurate segmentation even in areas with overlapping instances. Five categories of nuclei are predicted: nuclei of neoplastic, non-neoplastic epithelial, inflammatory, connective tissue and necrotic cells. For each image patch, HoVer-Net generates a set of nuclear centroids together with their cell type and the corresponding nuclear segmentation mask. Nuclear composition features in a patch can then be collected by counting the number of the five types of nuclei in it.

2.3.2. Nuclear morphological features (NMF)

Assuming that the morphology of different types of nuclei is associated with HER2 status, we extract 15 nuclear morphological features such as nuclear size, eccentricity, orientation, and length of the major axis (see Table S3) using the output binary mask of each nucleus. Therefore, each detected nucleus is represented by a 15-dimensional feature vector which contains 15 different morphological properties. We use the mean and standard deviation of the 15 feature values resulting in a 30-dimensional feature vector for each patch.

2.3.3. Neural embeddings (NE)

One of the strengths of deep neural networks is their ability to learn high-level features based on colour, frequency domain, edge detectors, texture and so on from image pixels. An image patch is fed into the network and transformed several times through convolutional layers in the network. During these transformations, the network is able to learn new and increasingly complex features of the input image. In this work, we experiment with two different types of neural embedding features. In order to extract a strong and representative set of features, the first neural embedding feature is obtained from the last convolutional layer of ResNet50 (He et al., 2016) due to its excellent performance in recent computer vision tasks. The model was trained on the ImageNet dataset (Deng et al., 2009) which is a large visual database designed for visual object recognition research. For the second neural embedding feature, we extracted a domain-specific 2048-dimensional representation from our in-house cellular composition prediction model called ALBRT (Dawood et al., 2021). ALBRT has been trained on the TCGA Breast cancer dataset (TCGA-BRCA) for predicting the counts of different types of cells in a given patch and is based on the Xception (Chollet, 2017) network with depthwise separable convolution along with self-supervised learning for rotational invariance.

2.3.4. DAB density estimates (DDE)

Areas of membranous DAB staining in IHC images can reveal the level of HER2 protein expression at a cellular level. Based on this, we utilised paired H&E and IHC images in the HER2C dataset to develop a deep convolutional neural network predictor to estimate the level of HER2 expression (DAB density estimates) in a given H&E image region which is then used as a node-level feature in the graph neural network. First, we performed affine registration on 5 H&E and IHC paired images from the HER2C dataset (Qaiser et al., 2018) by taking the H&E WSIs as the reference image and extracting several control points pairs at 40× resolution from each image pair. Fig. 3(a) shows an example H&E and IHC pair and the corresponding control points (yellow dots). The calculated affine transform matrix consisting of rotation, scaling and translation components is applied on all H&E image patches to get the corresponding IHC images. Fig. 3(a) gives three examples of the H&E and registered IHC images which shows high registration accuracy even at the highest resolution (40×). Second, in the registered IHC images, we convert their RGB colour space to Haematoxylin-Eosin-DAB (HED) colour space and calculate the percentage of DAB staining from the DAB channel.

In total, we collect more than 6000 H&E patches (size 512 × 512 pixels) and their corresponding DAB density values. Architecture of the proposed regression model is shown in Fig. 3(b). The feature extraction component of the network is inspired by ResNet50. Compared to the standard ResNet50 implementation, we add two fully connected layers after the feature extraction component with 2048 and 128 neurons, respectively. In order to evaluate the performance of our DAB-density regression model, we performed leave-one-WSI-out cross validation using the collected dataset and calculate the Pearson correlation coefficient (PCC) to measure the linear correlation between the ground truth and model prediction (Fig. 3(c)). Strongly positive correlation can be observed in all the 4 folds, achieving the averaged PCC 0.90 ($p < 0.0001$). Fig. 4 shows the visualisation of the estimated DAB density on a HER2-negative (first row) and a HER2-positive (second row) case, respectively. The WSIs shown in Fig. 4 are unseen by the trained model. It can be observed from the HER2-negative case that the majority of the tissue region have low estimated DAB density revealing the lack of HER2 protein expression. Compared to the negative case, the HER2-positive case has larger areas with high estimated DAB density as observed from the orange and red areas in the heatmap. The highlighted activation areas (zoomed-in) in the generated heatmap are consistent with the DAB density in the corresponding IHC images. This supports the idea of using DAB density as a potential feature for HER2 status prediction from H&E WSIs in computational pathology.

2.4. Adaptive spatial agglomerative clustering

As the number of patches in a WSI can be quite large, we group spatially neighbouring regions with high degree of similarity in the feature space in order to reduce the computational cost of downstream analysis. This is achieved using adaptive spatial agglomerative clustering which relies on a patch-level similarity kernel (see Algorithm 1). We use a feature space Gaussian kernel

Algorithm 1 WSI graph construction.

INPUT: A set of n patches in a given WSI - each represented by its spatial coordinates $\boldsymbol{g}_j$ and feature vector $\boldsymbol{h}_j$.
OUTPUT: Graph representation $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ of the WSI.
PARAMETERS:
$s_{\min} = 0.8$, $d_{\max} = 4000$ pixels at 0.25mpp
STEPS:
1: Perform agglomerative clustering using the similarity metric $k(\cdot,\cdot)$ with average linkage. Cluster agglomeration takes place up to a minimum similarity threshold of $s_{\min}$. For a given WSI X, this results in a cluster set $C(X)$ with each patch $x_j \equiv \left(\boldsymbol{g}_j, \boldsymbol{h}_j\right) \in X$ assigned to exactly one cluster $c \in C(X)$.
2: For each cluster set c, compute the geometric centre of all its constituent patches $\boldsymbol{p}_c = \frac{1}{|c|}\sum_{(\boldsymbol{g_j},\boldsymbol{h_j})\in c} \boldsymbol{g}_j$.
3: For each cluster set c, compute the aggregated feature vector of all its constituent patches $\boldsymbol{u}_c = \frac{1}{|c|}\sum_{(\boldsymbol{g_j},\boldsymbol{h_j})\in c} \boldsymbol{h}_j$.
4: Construct a vertex set V for the given WSI X with each cluster $c \in C(X)$ as a node represented by $\boldsymbol{v}_c \equiv (\boldsymbol{p}_c, \boldsymbol{u}_c)$.
5: Use Delauney triangulation to construct the edge set E based on the geometric coordinates of cluster centres with a maximum distance connectivity threshold of $d_{\max}$ pixels.

$k_h(\boldsymbol{h}_a, \boldsymbol{h}_b) = \exp(-\lambda_h \|\boldsymbol{h}_a - \boldsymbol{h}_b\|)$ and a geometric Gaussian kernel $k_g(\boldsymbol{g}_a, \boldsymbol{g}_b) = \exp\left(-\lambda_g \|\boldsymbol{g}_a - \boldsymbol{g}_b\|\right)$ to determine the degree of similarity between feature space representations and geometric coordinates of two patches in a given WSI (indexed by a and b), where λ_h and λ_g control the degree of similarity in feature space and distance in the geometric coordinates, respectively. We use a joint product kernel $k(x_a, x_b) = k_h(\boldsymbol{h}_a, \boldsymbol{h}_b) k_g(\boldsymbol{g}_a, \boldsymbol{g}_b)$ to model the overall similarity between two patches in a WSI. Note that the kernel product implies a tensor product of underlying features. The pairwise similarity matrix of all patches in a given WSI is then used as a similarity metric in Agglomerative clustering (Müllner, 2011) with average linkage. This is done such that cluster agglomeration takes place up to a minimum similarity threshold of $s_{\min}$. $s_{\min} \in [0, 1]$ affects the number of clusters in the WSI with large values of the parameter resulting in fewer and more heterogeneous clusters and vice-versa. Ideally, we would like to have small clusters to capture minor variations in the histological landscape. However, this has an impact on computational complexity of the model. In this work, we set $s_{\min} = 0.8$ to get a reasonable compromise between these two factors and evaluate the model performance with different $s_{\min}$ settings in Section 3.3. For a given WSI X, this results in a cluster set $C(X)$ with each patch $x_j \equiv \left(\boldsymbol{g}_j, \boldsymbol{h}_j\right) \in X$ assigned to exactly one cluster $c \in C(X)$. It is important to note that each cluster contains a number of patches and the number of clusters across all the WSIs can be different depending upon the sizes and morphological complexity of the WSIs.

2.5. Graph construction

We construct a WSI-level graph representation $G(X) \equiv (V, E)$ of a given WSI X based on its clusters (Sharma et al., 2015; Lu et al., 2019b; 2019a). The graph representation of a WSI X consists of a vertex set V with each cluster $c \in C(X)$ as a node represented by $\boldsymbol{v}_c \equiv (\boldsymbol{p}_c, \boldsymbol{u}_c)$ based on the geometric centre $\boldsymbol{p}_c = \frac{1}{|c|}\sum_{(\boldsymbol{g_j},\boldsymbol{h_j})\in c} \boldsymbol{g}_j$ and aggregated feature vector of all its constituent patches $\boldsymbol{u}_c = \frac{1}{|c|}\sum_{(\boldsymbol{g_j},\boldsymbol{h_j})\in c} \boldsymbol{h}_j$. The edge set $E \subseteq V \times V$ represents a finite set of edges between nodes. In order to capture signalling patterns between components of the tissue, the edge set is constructed by using Delauney triangulation based on the geometric coordinates of cluster centres with a maximum distance connectivity threshold of $d_{\max}$ pixels (Chew, 1989). This results in a planar graph, i.e., no two edges in the graph intersect each other. $d_{\max}$ controls the connectivity among clusters and is conceptually associated with the range of histological signaling. Larger $d_{\max}$ increases the number of connected neighbours of each cluster and the size of sub-graphs. It allows for progressively information sharing between neighbouring graph nodes. In this work, we choose $d_{\max}$ as 4000 pixels (at 0.25 microns per pixel) which ensures that most clusters have connected neighbours within 1 mm.

2.6. GNN architecture

The proposed architecture uses a Graph Neural Network (GNN) to generate both regional (node-level) and WSI-level predictions based on the above WSI-level graph representation. GNNs have the advantage of being inherently invariant to rotation and translation of graph nodes (Chami et al., 2020) and can learn progressively abstract node-level feature embdeddings across their layers through message passing or information sharing between neighbouring graph nodes. The architecture of the proposed graph neural network is shown in Fig. 5. It consists of multiple edge-convolution (EdgeConv) layers with each layer using a multi-layer perceptron (MLP) to produce a feature embedding of a node in the graph based on the feature embeddings of the node itself and its neighbouring nodes generated by previous layers (Wang et al., 2019). The first layer uses the original node-level features discussed above. The MLP in each layer uses the embedding of a node and the difference of its node embedding from its neighbours. Consequently, each GNN layer accumulates information from progressively higher order neighbours of each node. Mathematically, the output feature representation of an EdgeConv layer $l = 1 \ldots L$ in the GNN with L layers for a given node at index k in the input graph can be written as:

$$\boldsymbol{u}_k^{(l)} = \sum_{j \in N_k} H^{(l)}\left(\boldsymbol{u}_k^{(l-1)}, \boldsymbol{u}_j^{(l-1)} - \boldsymbol{u}_k^{(l-1)}; \theta_l\right). \tag{1}$$

In the above equation, $\boldsymbol{u}_k^{(0)} = \boldsymbol{u}_k$, N_k denotes the neighbourhood of node k and $H^{(l)}$ represents the multi-layer perceptron (MLP) with trainable weights θ_l. EdgeConv constructs a local neighbourhood graph centred at $\boldsymbol{u}_k$ and applies a convolution-like operation on the edge connecting the neighbourhood pair of node $\boldsymbol{u}_k$. As illustrated in Eq. (1), it can explicitly combine global shape structure captured by the features of node $\boldsymbol{u}_k$ with local neighbourhood information, captured by $\boldsymbol{u}_j - \boldsymbol{u}_k$ (Wang et al., 2019). As shown in Fig. 5, the feature embedding $\boldsymbol{u}_k^{(l)}$ of a node $\boldsymbol{v}_k \equiv (\boldsymbol{p}_k, \boldsymbol{u}_k) \in V$ from a GNN layer is passed to a corresponding linear layer to generate node level predictions $f_l(\boldsymbol{v}_k) = \boldsymbol{w}_l^T \boldsymbol{u}_k^{(l)}$. These node level prediction scores are then pooled to generate layer-wise WSI-level predictions, i.e., $F_l(\boldsymbol{G}) = \sum_{\forall v \in \boldsymbol{V}} f_l(v)$. These scores are then summed to produce the overall WSI-level prediction score, i.e., $F(\boldsymbol{G}; \theta) = \sum_{l=0}^{L} F_l(\boldsymbol{G})$ with all trainable parameters θ. It is important to note that SlideGraph$^+$ uses node-level aggregation as opposed to SlideGraph (Lu et al., 2020) which only has WSI-level output.

In this work, we use three EdgeConv layers ($L = 3$) which have 16, 16 and 8 neurons respectively in their MLP with a linear layer, followed by a batch normalisation (BN) and a Rectified Linear Unit activation (RELU) layers.

2.7. Loss function and hyperparameter settings

The proposed SlideGraph$^+$ is implemented using the PyTorch Geometric (PyG) library (Fey and Lenssen, 2019; Paszke et al., 2017). During the training, WSI-level prediction score is compared to the WSI-level target and weights in the EdgeConv layers and the base neural network block are updated through backpropagation. In each training batch, a set of positive cases B^+ and a set of negative cases B^- are chosen in a stratified manner. The loss function is

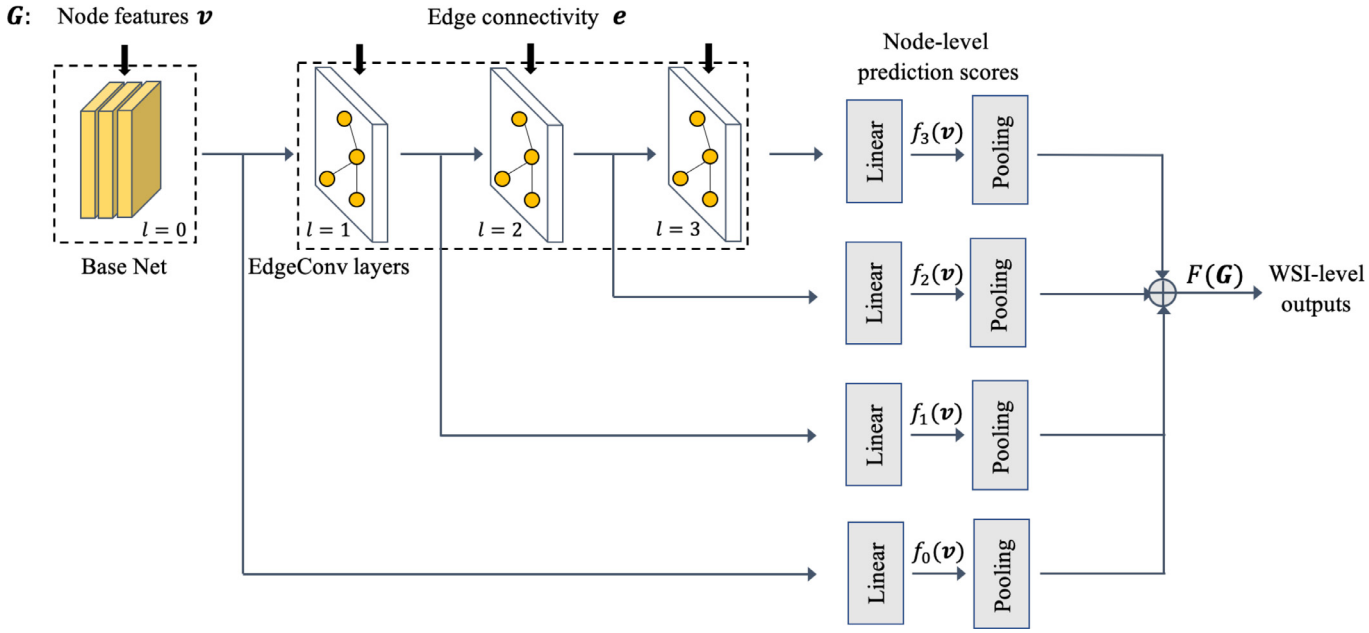

Fig. 5. Architecture of the proposed SlideGraph⁺ model for graph based WSI classification. The Base Net block is composed of a convolutional layer, Batch Normalisation and Rectified Linear Unit activation (RELU) layers. Graph neural network layer is structured with edge convolution (EdgeConv) layers whose mathematical expression is shown in (1). Pooling, Linear refer to the pooling and linear layers.

Table 1
A comparison of SlideGraph⁺ and the state-of-the-art methods using 5-fold cross validation on the TCGA-BRCA dataset. *: No standard deviation was reported in the original paper.

Method	Feature (dimension)	AUROC (mean ± std)
Campanella et al. (2019)	Resnet34	0.67 ± 0.05
Kather et al. (2019a)	Shufflenet	0.62*
Kather et al. (2019b)	Resnet18	0.68 ± 0.06
Rawat et al. (2020)	Fingerprints (512)	0.71*
Lu et al. (2021)	CLAM (1024)	0.64 ± 0.06
Li et al. (2021)	DSMIL (512)	0.65 ± 0.09
SlideGraph⁺	DAB density estimates (4)	0.75 ± 0.02

Table 2
A comparison of different features under SlideGraph⁺ architecture. For all the results shown, the training is done using the TCGA-BRCA dataset.

Method	Feature (dimension)	AUROC (mean ± std)
SlideGraph⁺ (5-fold cross validation on TCGA-BRCA dataset)	Nuclear composition (5)	0.71 ± 0.02
	Cellular morphology (30)	0.72 ± 0.05
	Embedding (Resnet50) (2048)	0.72 ± 0.07
	Embedding (ALBRT) (2048)	0.69 ± 0.04
	DAB density estimates (4)	0.75 ± 0.02
	Nuclear composition + Cellular morphology (35)	0.75 ± 0.04
	Nuclear composition + Cellular morphology + DAB (39)	0.75 ± 0.08

designed as a pairwise ranking based hinge-loss function with the mathematical formulation as follows:

$$\mathcal{L}(B^+, B^-; \theta) = \sum_{i \in B^+} \sum_{j \in B^-} \max\big(0, 1 - (F(G_i; \theta) - F(G_j; \theta))\big). \quad (2)$$

The minimisation of the loss function is implemented by using adaptive momentum-based optimisation (Kingma and Ba, 2014) with the learning rate 0.001 and a weight decay 0.0001. After training, the performance of the predictor is evaluated over test datasets. We use Area under Receiver Operator Characteristic (AUROC) curve and Precision-Recall (AUPR) curve to evaluate the predictive performance over test sets (Davis and Goadrich, 2006).

3. Results and discussion

3.1. HER2 status prediction

The current published state-of-the-art method by Rawat et al. (2020) gives AUROC values of 0.71 under five-fold cross-validation. In line with previous methods, we report AUROC for comparison. Table 1 shows the AUROC values by the proposed SlideGraph⁺, existing methods and state-of-the-art results. Using the same cross-validation strategy (Rawat et al., 2020), our proposed SlideGraph⁺ model with DAB density estimates achieves the best AUROC with 0.75 ± 0.02.

Performance of all proposed feature compositions under the SlideGraph⁺ framework are presented in Table 2. The Distributions of AUROC values achieved using different feature compositions are shown as box-plots in Fig. 6. It can be observed that most feature compositions under the SlideGraph⁺ framework achieve higher AUROC values than the state-of-the-art methods. Among all the features, feature combinations 'Nuclear composition + Nuclear morphology + DAB density estimates', 'Nuclear composition + Nuclear morphology' and 'DAB density estimates' exceed the state-of-the-art by a large margin, obtaining the maximum AUROC value of 0.75. The SlideGraph⁺ model with DAB density estimates achieves the smallest standard deviation with 0.02 in AUROC, proving its stability in HER2 status prediction. In addition, the estimated DAB density feature only has 4 dimensions, leading to the fewest number of training parameters and highest computational efficiency.

In order to compare the performance of the GNN in comparison to a naive aggregation of DAB density prediction scores, we average the estimated DAB densities from the four trained regression models on each patch and use three aggregating strategies (average pooling, max pooling and majority voting) to generate the overall WSI-level DAB density. Here we confine the average pooling and majority voting strategies on patches whose estimated DAB density is higher than 0.1. We calculate three type of DAB features – namely maximum DAB density estimates, majority DAB density es-

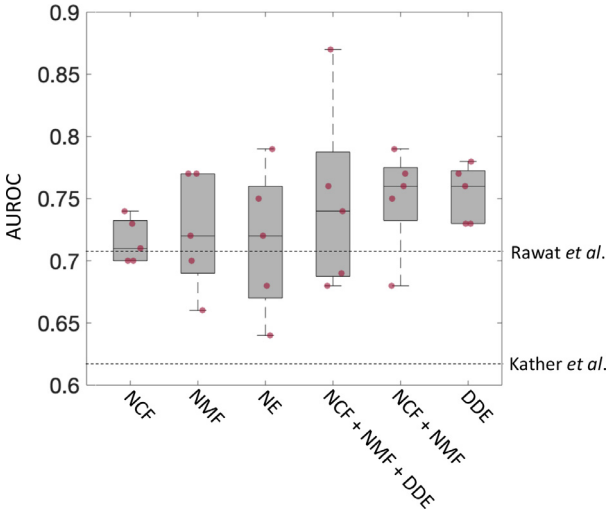

Fig. 6. AUROC values using different feature compositions during the five-fold cross validation on TCGA-BRCA. Dashed lines show the state-of-the-art AUROC values. NCF: Nuclear composition features; NMF: Nuclear morphological features; NE: Neural embeddings (Resnet50); DDE: DAB density estimates.

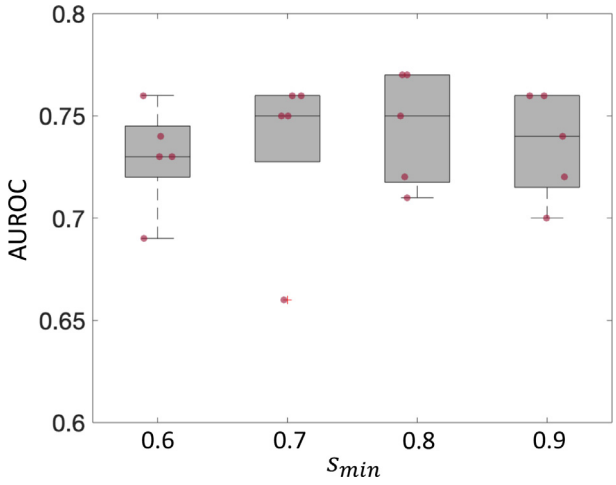

Fig. 8. AUROC values using different $s_{\min}$ values during the five-fold cross validation on TCGA-BRCA.

timates and average DAB density estimates respectively – on each WSI. As can be seen from Table S1, among all the three DAB density estimates features, the average DAB density estimates obtains the highest AUROC 0.598. The above results demonstrate the superiority of our proposed SlideGraph$^+$ architecture. Combining DAB density estimates with the help of a graph gives much higher AUROC and better HER2 prediction performance than the DAB density estimates on its own.

DAB density feature would be most helpful for further studies. DAB density feature represents the level of membranous DAB staining on IHC images which is widely used by pathologists to evaluate the level of HER2 protein expression and predict HER2 status. This feature can be applied on other receptor status prediction as long as corresponding DAB density estimating model is trained on the specific receptor data. In addition, the dimension of DAB density feature is small (4 in this paper) resulting in lower computational burden.

Nuclear composition and nuclear morphology capture cellular composition and morphological information from H&E images. However, these two features are collected from cell detection and classification results at the WSI-level which can be computationally expensive. In addition, the accuracy of the trained cell detection and classification model can have a significant impact on predictive performance of the WSI-level prediction task.

Embedding features learn high-level features based on colour, frequency domain, edge detectors, texture and so on from image pixels. However, this feature has very high dimension (1024 or 2048 in this paper) which significantly increases the computational complexity.

It is interesting to see that, between the two neural embedding features, feature collected from the model trained on breast cancer dataset does not give superior performance than the one on ImageNet dataset. It may due to the reason that the feature repre-

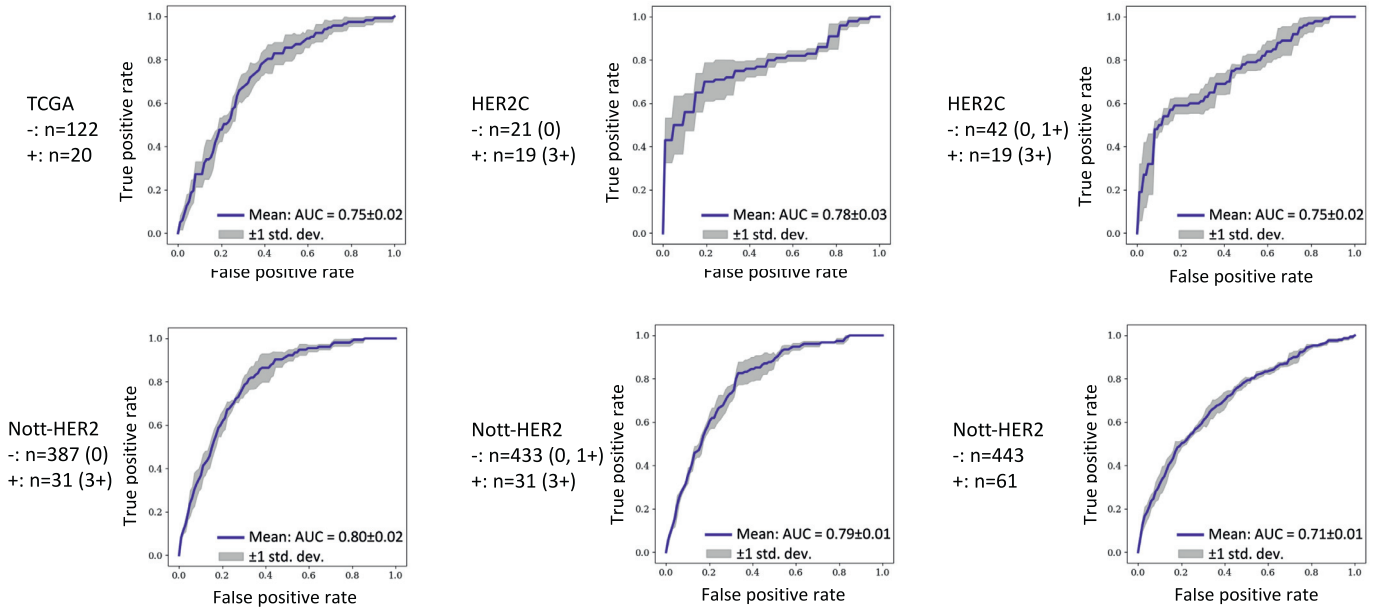

Fig. 7. ROC curves when testing the trained SlideGraph$^+$ classification model on the TCGA test dataset and on the other two independent test datasets (HER2C and Nott-HER2).

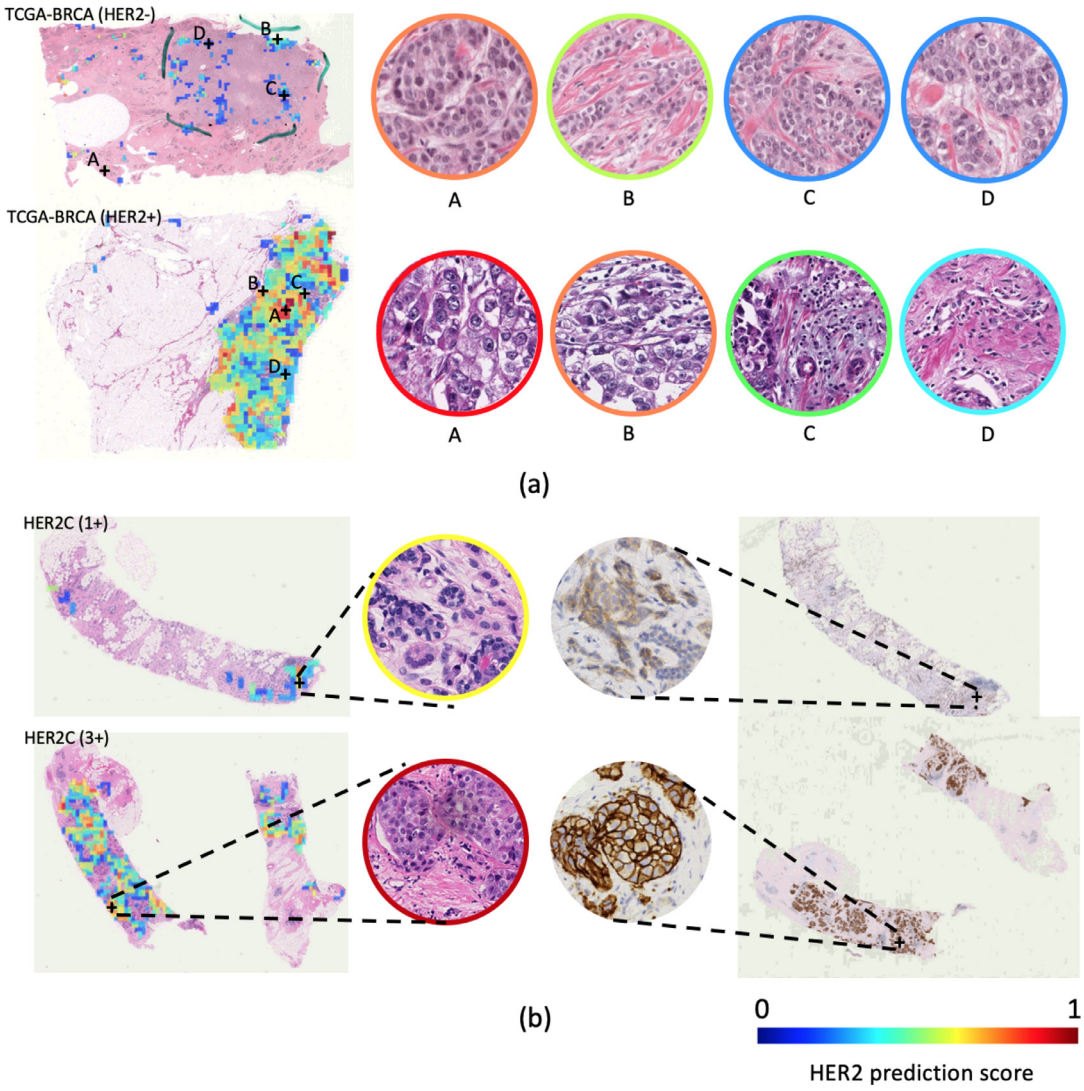

Fig. 9. Example heatmaps of node-level prediction scores: (a) cases from TCGA-BRCA; (b) cases from HER2C. Top row: HER2 negative; Bottom row: HER2 positive. Boundary colour of each zoomed-in region represents its contribution to HER2 positivity (prediction score). A-D in (a) denotes the positions of the zoomed-in regions. IHC images in (b) illustrate that the H&E regions with high HER2 prediction scores are consistent with the strongly stained DAB areas in the corresponding IHC images.

Table 3
External validation of SlideGraph$^+$ with DAB density estimates on datasets from multiple centres. (0 / 3+): differentiating negatives cases (status 0) and positive cases (3+); (0, 1+ / 3+): add 1+ cases to the negative group; (−/+): taking 2+ into consideration and differentiating all negative cases and positive cases. *: Baseline results are not available on these datasets.

Method	Test set	AUROC (mean ± std)
SlideGraph$^+$ (Independent validation on HER2C and Nott-HER2 datasets)*	HER2C (0 / 3+)	0.78 ± 0.03
	HER2C (0, 1+ / 3+)	0.75 ± 0.02
	Nott-HER2 (0 / 3+)	0.80 ± 0.02
	Nott-HER2 (0, 1+ / 3+)	0.79 ± 0.01
	Nott-HER2 (- / +)	0.71 ± 0.01

sentation from ALBRT model is problem specific (i.e cellular composition specific) but Resnet gives more generic features.

3.2. Independent validation

We then test our trained model on two independent test datasets: HER2C and Nott-HER2. Here, we utilise the model trained using DAB density estimates due to its superior performance and simplicity. As can be seen from Table 3, on HER2C dataset, the model achieves mean AUROC of 0.78 when differentiating negatives cases (status 0) and positive cases (3+). When we add 1+ cases to the negative group, our trained model achieves mean 0.75 AUROC value. For the Nott-HER2 dataset, our trained model achieves mean AUROC of 0.80 (0/3+), 0.79 (0, 1+/3+) and 0.71 (−/+) respectively. Corresponding ROC curves can be seen in Fig. 7. The independent validation on multi-centre datasets demonstrates the generalisation ability of the proposed SlideGraph$^+$ model.

3.3. Performance comparison with different $s_{\min}$ settings

Here, we choose $s_{\min}$ ranging from 0.6 to 0.9 to evaluate model's sensitivity on the number of clusters. We use DAB density feature for computational efficiency and do 5-fold cross validation on the TCGA-BRCA dataset. The distribution of AUROC using different $s_{\min}$ values are shown as box-plots in Fig. 8. It can be observed that the model performance remains stable and reaches the best average AUROC value (0.75) when $s_{\min}$ is 0.7 or 0.8. The average AUROC value decrease slightly (0.74) when $s_{\min}$ is 0.6 or 0.9. Therefore, setting $s_{\min}$ as 0.8 gives the most reasonable compromise between cluster homogeneity and computational complexity.

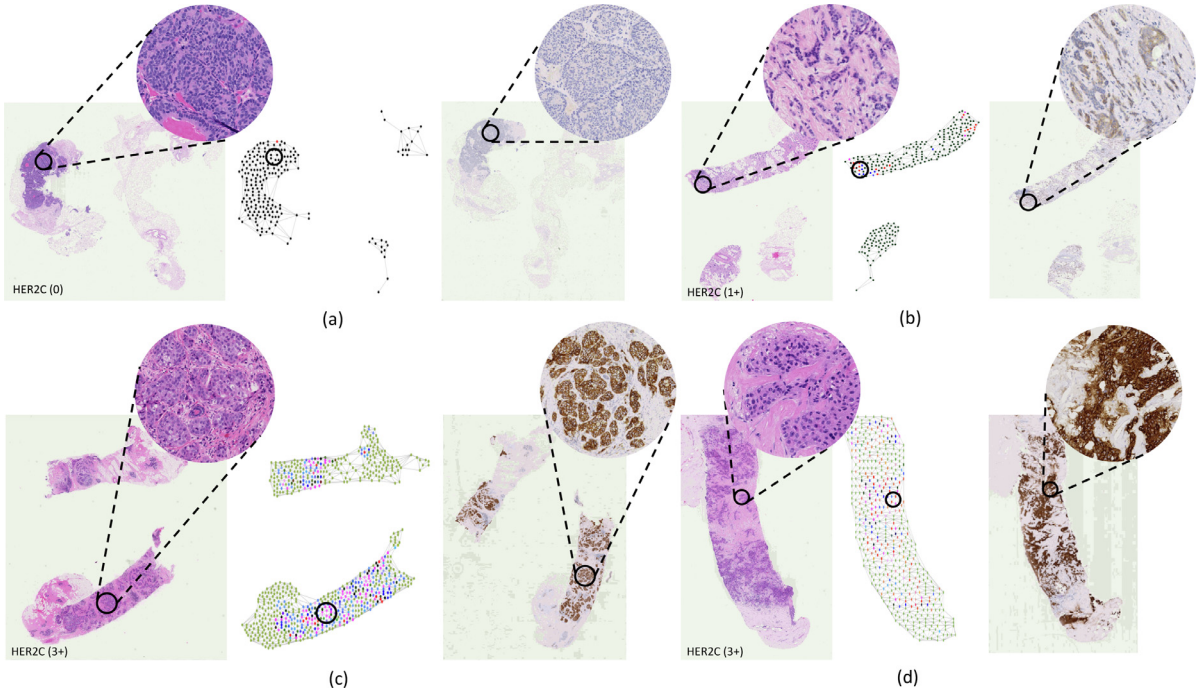

Fig. 10. Visualisation of node-level prediction on independent HER2C test dataset. Top row: HER2-negative; Bottom row: HER2-positive.

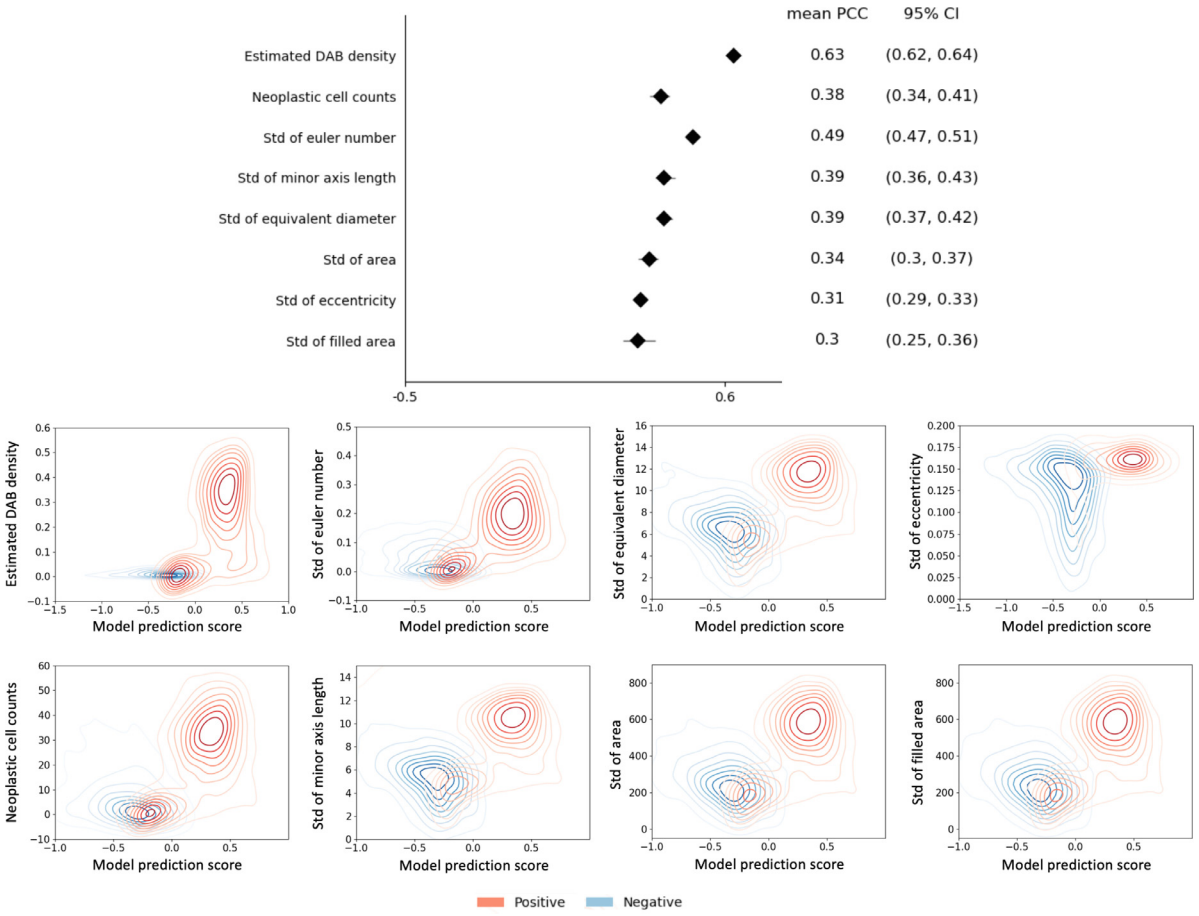

Fig. 11. Top: Correlation between nuclear pleomorphism and model prediction scores for ten patients from TCGA-BRCA cohort. Each row gives the mean and confidence interval (CI) of PCC between the graph node-level prediction score and a specific feature. For all features, $p < 0.0001$. Bottom: Density plots of nuclear pleomorphism related features and model prediction scores between positive cases (red) and negative cases (blue).

This experiment proves that the performance of the model is stable and not sensitive to the hyperparameter $s_{\min}$.

3.4. Performance comparison using AUPR

Despite significant class imbalance, previously published works (Kather et al., 2019a; Rawat et al., 2020) did not report prediction results in terms of area under the precision-recall curve (AUPR). In this paper, in addition to AUROC, we also record the AUPR results in each experimental setting and show the values in the supplementary Table S2 and Fig. S1. In five-fold cross-validation on the TCGA-BRCA dataset, the SlideGraph$^+$ model with DAB density estimates achieves the best AUPR 0.37 $\pm$ 0.03. In comparison, the DAB density estimate feature without the use of the proposed graph network achieves a maximum AUPR value of only 0.20. This shows the significant impact of our proposed SlideGraph$^+$ architecture. We also test our trained model on the other two independent test datasets. On the HER2C dataset, the model achieves mean AUPR 0.82 when differentiating definitive negative cases (status 0) and positive cases (3+). When we add 1+ cases to the negative group, our trained model achieves mean AUPR value of 0.64. For the Nott-HER2 dataset, our model achieves mean AUPR 0.25 (0/3+), 0.20 (0, 1+/3+) and 0.28 ($-/+$) respectively. It is important to note that AUPR in the independent test sets are not comparable across datasets because the number of cases in both classes and the ratio between them varies across different settings.

3.5. Visualisation of HER2 predictions

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, let us examine the node-level prediction performance on several cases from TCGA-BRCA and HER2C datasets. Fig. 9(a) shows the overlay of heatmaps and four zoomed-in regions which have different levels of HER2 prediction score. It can be observed that only a few areas in the negative sample contribute to the HER2 positivity while majority of the tissue regions in the positive case have high HER2 prediction scores. Same can be observed on sample images from the HER2C dataset in Fig. 9(b). It should also be noted that regions with high HER2 prediction scores are consistent with high DAB intensity areas in the corresponding IHC images.

We then convert the node-level prediction score into a false colour representation of each node. This results in a WSI-level graph visualisation in which the colour of each node is based on its node-level prediction score. Fig. 10 shows the results of this visualisation for two HER2-negative (top row) and two HER2-positive (bottom row) WSIs. One can observe clear differences in the graphs of the two classes: note the prevalence of red and blue areas in HER2-positive WSIs and dark green areas in HER2-negative WSIs. This supports the overall idea of using WSI-level graphs proposed in this work and the utility of incorporating global context for machine learning problems in computational pathology.

3.6. Correlation between nuclear pleomorphism and model prediction score

We conduct further analysis on nuclei pleomorphism related features that contribute to the HER2 prediction. We include five HER2-positive and five HER2 negative-cases in this experiment and calculate the Pearson correlation coefficient (PCC) between the node-level prediction score and cell nuclei pleomorphism related features. In Fig. 11 (top), we show the mean, confidence interval (CI) of PCC and limit our discussion to features whose mean PCC is above 0.3. We can see that the estimated DAB density feature gives the highest PCC value of 0.63 (95% CI 0.62, 0.64). Among the nuclear composition features, neoplastic cell counts contribute more to HER2 positivity prediction, with mean PCC 0.38 and 95% CI (0.34, 0.41). This is plausible because regions with HER2 overexpression normally have larger number of neoplastic cells.

Among all the nuclear pleomorphism related features, standard deviation of the Euler number gives the strongest positive correlation value of 0.49 (95% CI 0.47–0.51, $p < 0.0001$). Mathematically, in a 2D nuclear mask, the Euler number is the number of objects minus the number of holes. Hence, the standard deviation of the Euler number may capture the diversity of nuclear morphology and chromatin texture. Higher values represent major morphological differences. Standard deviation values of minor axis length, equivalent diameter, area, eccentricity and filled area are another five nuclear pleomorphism related features which give mean PCC above 0.3 with $p < 0.0001$. These five features are associated with the significance of variation in shape and sizes of the cells. Density plots of features and model prediction scores between positive cases (red) and negative cases (blue) are shown in Fig. 11 (bottom). Clear separation can be observed between the positive and negative groups. The observations here point to the association of nuclear pleomorphism with HER2 positivity and cancer progression.

3.7. Comparison of computational efficiency

We have also compared the computational efficiency of patch-based and the proposed SlideGraph$^+$ model using a single Nvidia Titan RTX GPU. Once the patches and graphs are obtained from the WSI, the average single-fold training time for the baseline model (Kather et al., 2019b) is 5.3 h and the testing time for a WSI is 1.2 s from patches to the final prediction. In comparison, SlideGraph$^+$ training for a single fold takes 2 min on average and 0.4 ms to get the label prediction from a single graph. In terms of the time of feature extraction on TCGA-BRCA dataset in which majority of the slides are tissue regions, extracting HoVer-Net related features (NCF and NMF) takes 29 min per slide on average. Neural embedding features (NE) takes 8.8 min while DAB density estimates (DDE) needs 4 min to process each slide on average. As patch level features only need to be extracted once, the graph-based modelling approach can save large amounts of time when we fine-tune predictive model and run the training process multiple times for estimation of predictive performance. In addition, extracted features can be used for other prediction tasks directly while other baseline methods need to be trained on patches for each task.

4. Conclusions

In this paper, we proposed SlideGraph$^+$ as a generic method that couples WSI-level graph representation with a graph neural network for capturing the global context of a WSI and showed its effectiveness for prediction of HER2 status directly from WSIs of H&E stained BCa tissue slides. 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 SlideGraph$^+$ can effectively incorporate both cell-level and contextual information by using different feature compositions and graph convolution. We also proposed a DAB density regression model which can predict HER2 specific DAB density directly from H&E images. Experimental results for clinically important tasks of HER2 status prediction show that the proposed SlideGraph$^+$ method with estimated DAB density feature can produce higher accuracy than the state-of-the-art techniques. SlideGraph$^+$ can also be applied to other problems in computational pathology, such as recurrence and survival prediction, anti-HER2 treatment efficacy prediction.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Wenqi Lu: Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Michael Toss:** Data curation, Resources. **Emad Rakha:** Data curation, Resources. **Nasir Rajpoot:** Funding acquisition, Project administration, Supervision. **Fayyaz Minhas:** Conceptualization, Methodology, Project administration, Supervision, Writing – original draft.

Acknowledgements

We would like to gratefully acknowledge all subjects whose whole slide image data has been used in the development of this study. This paper is supported by the PathLAKE Centre of Excellence for digital pathology and artificial intelligence which is funded from the Data to Early Diagnosis and Precision Medicine strand of the governments Industrial Strategy Challenge Fund, managed and delivered by Innovate UK on behalf of UK Research and Innovation (UKRI). PathLAKE is a Research Ethics Committee (REC) approved research database, reference 19/SC/0363.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.media.2022.102486

References

Acs, B., Rantalainen, M., Hartman, J., 2020. Artificial intelligence as the next step towards precision pathology. J. Intern. Med. 288 (1), 62–81.
Ahmad, A., 2019. Breast Cancer Metastasis and Drug Resistance: Challenges and Progress, vol. 1152. Springer.
Ali, S., Veltri, R., Epstein, J.A., Christudass, C., Madabhushi, A., 2013. Cell cluster graph for prediction of biochemical recurrence in prostate cancer patients from tissue microarrays. In: Medical Imaging 2013: Digital Pathology, vol. 8676. International Society for Optics and Photonics, p. 86760H.
Andrews, S., Hofmann, T., Tsochantaridis, I., 2002. Multiple instance learning with generalized support vector machines. In: AAAI/IAAI, pp. 943–944.
Bandi, P., Geessink, O., Manson, Q., Van Dijk, M., Balkenhol, M., Hermsen, M., Bejnordi, B.E., Lee, B., Paeng, K., Zhong, A., et al., 2018. 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.
Campanella, G., Hanna, M.G., Geneslaw, L., Miraflor, A., Silva, V.W.K., Busam, K.J., Brogi, E., Reuter, V.E., Klimstra, D.S., Fuchs, T.J., 2019. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat. Med. 25 (8), 1301–1309.
Chami, I., Abu-El-Haija, S., Perozzi, B., Ré, C., Murphy, K., 2020. Machine learning on graphs: a model and comprehensive taxonomy. arXiv preprint arXiv:2005.03675
Chew, L.P., 1989. Constrained Delaunay triangulations. Algorithmica 4 (1–4), 97–108.
Chollet, F., 2017. Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251–1258.
Cruz-Roa, A., Basavanhally, A., González, F., Gilmore, H., Feldman, M., Ganesan, S., Shih, N., Tomaszewski, J., Madabhushi, A., 2014. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In: Medical Imaging 2014: Digital Pathology, vol. 9041. International Society for Optics and Photonics, p. 904103.
Davis, J., Goadrich, M., 2006. The relationship between precision-recall and ROC-curves. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 233–240.
Dawood, M., Branson, K., Rajpoot, N.M., Minhas, F., 2021. ALBRT: cellular composition prediction in routine histology images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 664–673.
Demir, C., Gultekin, S.H., Yener, B., 2005. Augmented cell-graphs for automated cancer diagnosis. Bioinformatics 21 (suppl_2), ii7–ii12.
Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Ieee, pp. 248–255.
Fey, M., Lenssen, J. E., 2019. Fast graph representation learning with pytorch geometric. arXiv preprint arXiv:1903.02428
Farahmand, S., Fernandez, A.I., Ahmed, F.S., Rimm, D.L., Chuang, J.H., Reisenbichler, E., Zarringhalam, K., 2022. Deep learning trained on hematoxylin and eosin tumor region of interest predicts HER2 status and trastuzumab treatment response in HER2+ breast cancer. Mod. Pathol. 35 (1), 44–51.
Gamper, J., Koohbanani, N.A., Benet, K., Khuram, A., Rajpoot, N., 2019. Pannuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification. In: European Congress on Digital Pathology. Springer, pp. 11–19.
Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N., 2019. Hover-net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med. Image Anal. 58, 101563.
Gunduz, C., Yener, B., Gultekin, S.H., 2004. The cell graphs of cancer. Bioinformatics 20 (suppl_1), i145–i151.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
Hou, L., Samaras, D., Kurc, T.M., Gao, Y., Davis, J.E., Saltz, J.H., 2016. Patch-based convolutional neural network for whole slide tissue image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2424–2433.
Janowczyk, A., Madabhushi, A., 2016. Deep learning for digital pathology image analysis: acomprehensive tutorial with selected use cases. J. Pathol. Inform. 7.
Jaume, G., Pati, P., Anklin, V., Foncubierta, A., Gabrani, M., 2021. Histocartography: a toolkit for graph analytics in digital pathology. In: MICCAI Workshop on Computational Pathology. PMLR, pp. 117–128.
Kather, J. N., Heij, L. R., Grabsch, H. I., Kooreman, L. F., Loeffler, C., Echle, A., Krause, J., Muti, H. S., Niehues, J. M., Sommer, K. A., et al., 2019a. Pan-cancer image-based detection of clinically actionable genetic alterations. bioRxiv, 833756.
Kather, J.N., Pearson, A.T., Halama, N., Jäger, D., Krause, J., Loosen, S.H., Marx, A., Boor, P., Tacke, F., Neumann, U.P., et al., 2019. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat. Med. 25 (7), 1054–1056.
Kingma, D. P., Ba, J., 2014. Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980
LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324.
Li, B., Li, Y., Eliceiri, K.W., 2021. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14318–14328.
Lu, C., Wang, X., Prasanna, P., Corredor, G., Sedor, G., Bera, K., Velcheti, V., Madabhushi, A., 2018. 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. Springer, pp. 407–416.
Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F., 2021. Data-efficient and weakly supervised computational pathology on whole-slide images. Nat. Biomed. Eng. 5 (6), 555–570.
Lu, W., Duan, J., Orive-Miguel, D., Herve, L., Styles, I.B., 2019. Graph-and finite element-based total variation models for the inverse problem in diffuse optical tomography. Biomed. Opt. Express 10 (6), 2684–2707.
Lu, W., Duan, J., Veesa, J.D., Styles, I.B., 2019. New nonlocal forward model for diffuse optical tomography. Biomed. Opt. Express 10 (12), 6227–6241.
Lu, W., Graham, S., Bilal, M., Rajpoot, N., Minhas, F., 2020. Capturing cellular topology in multi-gigapixel pathology images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 260–261.
Müllner, D., 2011. Modern hierarchical, agglomerative clustering algorithms. arXiv preprint arXiv:1109.2378
Nahta, R., Yu, D., Hung, M.-C., Hortobagyi, G.N., Esteva, F.J., 2006. Mechanisms of disease: understanding resistance to HER2-targeted therapy in human breast cancer. Nat. Clin. Pract. Oncol. 3 (5), 269–280.
Network, C.G.A., et al., 2012. Comprehensive molecular portraits of human breast tumours. Nature 490 (7418), 61.
Nguyen, M.H., Torresani, L., De La Torre, F., Rother, C., 2009. Weakly supervised discriminative localization and classification: a joint learning process. In: 2009 IEEE 12th International Conference on Computer Vision. IEEE, pp. 1925–1932.
Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A., 2017. Automatic differentiation in pytorch.
Prewitt, J.M., 1979. Graphs and grammars for histology: an introduction. In: Proceedings of the Annual Symposium on Computer Application in Medical Care. American Medical Informatics Association, p. 18.
Qaiser, T., Mukherjee, A., Reddy Pb, C., Munugoti, S.D., Tallam, V., Pitkäaho, T., Lehtimäki, T., Naughton, T., Berseth, M., Pedraza, A., et al., 2018. 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.
Rawat, R.R., Ortega, I., Roy, P., Sha, F., Shibata, D., Ruderman, D., Agus, D.B., 2020. Deep learned tissue "fingerprints" classify breast cancers by ER/PR/Her2 status from H&E images. Sci. Rep. 10 (1), 1–13.
Ross, J.S., Slodkowska, E.A., Symmans, W.F., Pusztai, L., Ravdin, P.M., Hortobagyi, G.N., 2009. The HER-2 receptor and breast cancer: ten years of targeted anti–HER-2 therapy and personalized medicine. Oncologist 14 (4), 320–368.
Sharma, H., Zerbe, N., Lohmann, S., Kayser, K., Hellwich, O., Hufnagl, P., 2015. A review of graph-based methods for image analysis in digital histopathology. Diagn. Pathol. 1 (1), 1–51.
Sirinukunwattana, K., Snead, D., Epstein, D., Aftab, Z., Mujeeb, I., Tsang, Y.W., Cree, I., Rajpoot, N., 2018. Novel digital signatures of tissue phenotypes for predicting distant metastasis in colorectal cancer. Sci. Rep. 8 (1), 1–13.

Slamon, D.J., Clark, G.M., Wong, S.G., Levin, W.J., Ullrich, A., McGuire, W.L., 1987. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science 235 (4785), 177–182.
Tizhoosh, H.R., Pantanowitz, L., 2018. Artificial intelligence and digital pathology: challenges and opportunities. J. Pathol. Inform. 9.
Vahadane, A., Peng, T., Sethi, A., Albarqouni, S., Wang, L., Baust, M., Steiger, K., Schlitter, A.M., Esposito, I., Navab, N., 2016. Structure-preserving color normalization and sparse stain separation for histological images. IEEE Trans. Med. Imaging 35 (8), 1962–1971.
Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M., 2019. Dynamic graph CNN for learning on point clouds. ACM Trans. Graph. (TOG) 38 (5), 1–12.
Weyn, B., van de Wouwer, G., Kumar-Singh, S., van Daele, A., Scheunders, P., Van Marck, E., Jacob, W., 1999. Computer-assisted differential diagnosis of malignant mesothelioma based on syntactic structure analysis. Cytometry 35 (1), 23–29.
Whitney, J., Corredor, G., Janowczyk, A., Ganesan, S., Doyle, S., Tomaszewski, J., Feldman, M., Gilmore, H., Madabhushi, A., 2018. Quantitative nuclear histomorphometry predicts oncotype DX risk categories for early stage ER+ breast cancer. BMC Cancer 18 (1), 610.
Wolff, A.C., Hammond, M.E.H., Allison, K.H., Harvey, B.E., Mangu, P.B., Bartlett, J.M., Bilous, M., Ellis, I.O., Fitzgibbons, P., Hanna, W., et al., 2018. Human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical oncology/college of american pathologists clinical practice guideline focused update. Arch. Pathol. Lab. Med. 142 (11), 1364–1382.
Yarden, Y., 2001. Biology of HER2 and its importance in breast cancer. Oncology 61 (Suppl. 2), 1–13.
Yener, B., 2016. Cell-graphs: image-driven modeling of structure-function relationship. Commun. ACM 60 (1), 74–84.
Zhang, C., Platt, J.C., Viola, P.A., 2006. Multiple instance boosting for object detection. In: Advances in Neural Information Processing Systems, pp. 1417–1424.
Zhang, Q., Goldman, S.A., 2002. EM-DD: an improved multiple-instance learning technique. In: Advances in Neural Information Processing Systems, pp. 1073–1080.