Ph.D. Dissertation of Engineering

PROGNOSIS PREDICTION OF CANCER PATIENTS AND DISCOVERY OF CANCER MICROENVIRONMENT BIOMARKERS USING GRAPH NEURAL NETWORK

그래프 딥러닝을 활용한 암환자의 예후 예측

및 암 미세환경 마커 발굴

AUGUST, 2022

DEPARTMENT OF ELECTRICAL AND

COMPUTER ENGINEERING

SEOUL NATIONAL UNIVERSITY

YONGJU LEE

PROGNOSIS PREDICTION OF CANCER PATIENTS AND DISCOVERY OF CANCER MICROENVIRONMENT BIOMARKERS USING GRAPH NEURAL NETWORK

지도 교수 권 성 훈

이 논문을 공학박사 학위논문으로 제출함
2022 년 8 월

서울대학교 대학원
전기정보공학부
이 용 주

이용주의 공학박사 학위논문을 인준함
2022 년 8 월

위 원 장 _____ 김 선 _____ (인)

부위원장 _____ 권 성 훈 _____ (인)

위   원 _____ 이 정 우 _____ (인)

위   원 _____ 강 유 _____ (인)

위   원 _____ 명 재 경 _____ (인)

Abstract

In this dissertation, I propose a new approach named tumor environment-associated context analysis using deep graph learning (TEA-graph) based on a memory-efficient graph representation of WSIs and interpretable graph neural networks (GNNs) that considers the contextual features of the tumor environment in a semi-supervised manner. Gigapixel resolution whole-slide image (WSI), which is a digitally scanned tumor section, has changed the approach of pathological analysis in a quantitative and data-driven manner. The combination of deep learning and WSI brought more synergetic effects and revolutionized pathological analysis in clinically important tasks. Although deep learning on WSIs shows revolutionary results as an assistive tool for computer-aided diagnosis (CAD) and automated analysis, it still cannot account for clinically important pathological features such as the tumor microenvironment (TME), which is a "contextual pathological feature".

The TEA-graph extracts the "contextual pathological features" of WSIs in a memory-efficient and semi-supervised manner. While similar concepts were introduced that using graph neural networks for contextual learning and compressed the whole-slide image (WSI) in various ways, none of them introduce geometry structure features (positional features) which are important features to differentiate the various tumor microenvironment. The proposed TEA-graph utilized the WSI at once without losing contextual information on the WSI through a network compression technology named superpatch. Therefore, the TEA-graph is scalable to massive WSIs without manual annotation of experts and interpretable to researchers and clinicians. I validated the prognosis prediction performance of TEA-graph using data from a total of 4,967 patients consisting of four different types of cancer.

Especially, TEA-graph captured heterogeneous contextual features of clear cell renal cell carcinoma (ccRCC) on a dataset from 1366 patients and reflected these features to predict the prognosis of ccRCC. Additionally, TEA-graph provided interpretable contextual features and enabled us to categorize the contextual features into several groups that had different prognostic effects. Proposed results indicated that TEA-graph identified the complex interaction of histopathological features and therefore provided prognostic biomarkers in a data-driven manner. I showed the pathological features related to the favorable or unfavorable prognosis and suggested the angiogenetic pathological feature as a novel indicator of renal cell carcinoma. Furthermore, I checked which part of TEA-graph is critical for optimal performance with thorough hyperparameter screening and ablation studies, including performance comparison. Overall, the proposed method, including geometrical features, attention-based model, and the compressed network construction method, is critical in achieving optimal performance. Because contextual feature learning will introduce new insights to researchers and clinicians and overall performance increases to models, I believe that follow-up studies that incorporate other datasets, such as genetic materials or extend beyond pathology (radiology or MRI, CT), are possible. Overall, proposed TEA-graph can aid researchers from various fields that use deep-learning to the medical application.

Table of Contents

iii

List of Tables

List of Figures

Chapter 1.

Introduction

Histopathological biomarkers have been used for standard diagnosis tool to guide the medical practitioners and patients for managing the disease specifically tumors for decades. Recently, histopathological studies are changed into the quantitative analysis and machine learning or deep learning based methodologies are actively developed for various purpose. In this chapter, the utility of histopathological biomarker for tumor prognosis is introduced. Especially, tumor architectural features of renal carcinoma are described as the example of pathological biomarker. And the advantage of computational pathology which is digitized pathology image and utilization of data-based pathology method will be discussed. Also, recent trend of deep learning for computational pathology is briefly introduced and detailed description of how to predict the prognosis of patients using pathology image will be discussed.

1.1. Histopathological biomarker for tumor prognosis

1.1.1. Pathological evidence for prognosis

Histopathology refers to the examination of tissue under a microscope to study signs of disease. Histopathology in clinical medicine refers to the biopsy or surgical examination of the specimen by a pathologist after processing the specimen and placing the histological section on a glass slide [1]. Even recently, histopathology remains a manual process in which pathologists examine glass slides using conventional brightfield microscopy. In particular, the histopathology of tumors reflects the heterogeneous nature of the tumor environment, which is an important prognostic biomarker to evaluate the status of the tumor and guide treatment options. Therefore, histopathological analysis of tumor tissue has traditionally been used as the basis for diagnosis and prognosis. For example, the International Society of Urological Pathology (ISUP) grading system for renal cell carcinoma (RCC) concluded that tumor morphology, tumor grade, fuhrman nuclear grade, and sarcoma structure are strong prognostic indicators informing patient management [2]. In addition, the Gleason grading of prostate adenocarcinoma is a widely used grading system based on the histological pattern of the carcinoma cell arrangement in hematoxylin and eosin (H&E) stained sections, recommended for routine reporting of prognostic factors [3].

1.1.2. Tumor architecture features at renal carcinoma

Throughout the thesis, I focused on the pathological patterns of major subtype of renal cell carcinoma which is clear cell renal cell carcinoma (ccRCC). Renal cell carcinoma is well known for its histopathologic patterns related to prognosis, named

2

the World Health Organization/International Society of Urological Pathology (WHO/ISUP) or Fuhrman grade [4]. In addition, ccRCC has a heterogeneous morphological structure, which is suspected to be related to the genomic heterogeneity of ccRCC and its impact prognosis (Figure 1.1). Sarcomatous component, lymphatic invasion are important morphological structures in high-risk populations. The above-mentioned features can be measured via H&E stained tissue, a standard staining method for histopathological evaluation [5].
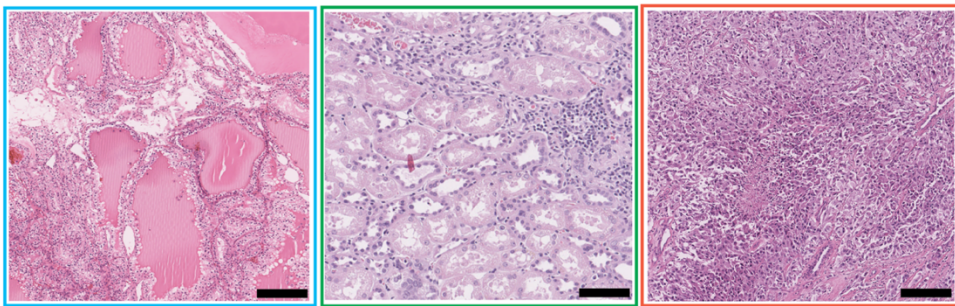
Figure 1.1 Example of pathological architecture patterns in ccRCC. Cystic change (blue); Papillary growth (green); Lymphocytic infiltration (red). Scale bar, 200μm.

1.2. Computational pathology for decision making

1.2.1. Computational pathology

The advent of high-resolution whole slide images (WSI) using digital slide scanners has changed the way histological analysis is performed. The development of brightfield and fluorescence slide scanners has made possible the virtualization and digitization of whole glass slides. A variety of full slide scanners with brightfield and fluorescence capabilities have been shipped with various hardware and software specifications. When histological glass slides are digitized, they can be viewed remotely by a pathologist on a computer screen or analyzed digitally using image analysis techniques. WSI has several unique characteristics that make it different

from other biomedical imaging data. First, WSI is a very high resolution image. Typically, full slide scanners have resolutions of 0.5 µm/pixel (20x resolution) and 0.25 µm/pixel (40x resolution). So, using a standard pathology slide (25 x 50 mm), the total pixel size is 100,000 x 200,000 pixels, which is 20 gigapixels. Second, WSI stands for multi-resolution image. A general configuration of the WSI of a full slide scanner, thought of as a pyramid of image data. WSI consists of multiple images of different resolutions. The basis of the pyramid is the highest resolution image data. And you can create a thumbnail image as a lower resolution version of the image so you can see the whole image at once [6].

1.2.2. Early attempt to apply machine learning for pathology

Computational pathology has emerged that leverages the wealth of information in WSI to alleviate the labor-intensive efforts of pathologists or analytical tools that can intuit the pathological features of a specimen through automating the quantification of pathological domains. For example, early efforts relied on the sorting of individual cells inside tissues and quantification of those cells using supportive vector machines for systematic analysis of the tumor milieu for breast cancer with genomic characteristics [7]. Others have measured quantitative features, including relational and global image features, in breast cancer epithelium and stroma. They used simple linear regression to classify epithelial and stromal regions and performed further nuclear-level analysis. The quantified features predicted the patient's prognosis [8]. All these early attempts are pioneering and bring clinically meaningful results. However, they relied on the classic machine learning tools to classify the pathological features inside the WSI which can't provide the sufficient performance.

1.3. Deep learning based computational pathology

1.3.1. Deep learning based pathological analysis

Recent developments in deep learning advanced computational pathology thanks to the superior performance of deep learning in image analysis. Initially, simple classification tasks such as classifying tumor tissue from normal tissue or subtyping within a single tumor were performed. Today, the classification and classification of different cell types in WSI is highly accurate and it is possible to predict the origin of oncogenic variants, gene expression, or metastases. In particular, deep-learning-based computational pathology has shown promising results in predicting microsatellite instability, an important molecular marker for prognosis and therapy, resulting from genomic differences between tumors. Also, unlike traditional data-driven pattern analysis, deep learning does not require manual feature extraction, so deep learning can extract useful features to maximize performance, and the extracted features can provide insight to pathologists. In that way, leveraging prognostic image markers from the vast amount of WSI possible with the collaboration of classical survival analysis and deep learning methods will allow pathologists to review existing biomarkers and provide intuition for prognosis-related pathological characterization.

1.3.2. Deep learning based time-to-event analysis

The purpose of survival analysis is to establish a link between covariates and event times. Survival analysis is a kind of regression problem, but it differs from traditional regression problems because only part of the training data is partially observable, which are censored. For censored subjects, no deaths can be observed, so the type of

censorship is called right censorship. The challenge in survival analysis is how to properly handle censored data. In general, survival time is modeled as a continuous non-negative random variable T from which I can derive the basic quantity, survival function, and hazard function for time-to-event analysis. The survival function $S(t)$ returns the probability of surviving after time t $S(t) = P(T > t)$. The hazard function $h(t)$ represents the approximate probability that an event will occur in small time intervals Δt.

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} \geq 0$$

There is a non-parametric survival analysis such as Kaplan-Meier analysis is also interesting method, however I will focus on the parametric analysis to estimate the analysis into deep learning method. By far the most popular model for learning from censored survival data is Cox's proportional hazards model. This method models the hazard function of the individual conditioned on the feature vector.

$$h(t_i) = h(t \mid x_{i1} \dots x_{ip}) = h_0(t) \exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right)$$

Cox's proportional hazards model is fitted by maximizing a partial likelihood function based on probability. Let $R_i = \{j \mid y_j \geq y_i\}$ is the set of subjects who remained event-free before time point y_i. If the subject experiences event at y_i, that subject must have the higher probability than the others who does not experience the event, which mean event-free up to y_i. It can be defined as follows

$P(subject\ experiences\ event\ at\ y_i \mid one\ event\ at\ y_i)$

$$= \frac{P(subject\ experiencecs\ event\ at\ y_i \mid event - free\ up\ to\ y_i)}{P(one\ event\ at\ y_i \mid event - free\ up\ to\ y_i)}$$

6

$$= \frac{h(y_i|x_i)}{\sum_{j=1}^{n} I(y_j \geq y_i) h(y_j|x_j)} = \frac{h_0(y_i) \exp(x_i^T \beta)}{\sum_{j=1}^{n} I(y_j \geq y_i) h_0(y_j) \exp(x_j^T \beta)}$$

$$= \frac{\exp(x_i^T \beta)}{\sum_{j \in R_i} \exp(x_j^T \beta)}$$

By setting the loss function as a partial likelihood function, the neural network model can function as a nonlinear cox regression model, which was first proposed by Faraggi and Simon by a single hidden layer multilayer perceptron [9].

$$\hat{\beta} = \underset{\beta}{\mathrm{argmax}} \, logPL(\beta) = \sum_{i=1}^{n} \delta_i \left[x_i^T \beta - \log\left(\sum_{j \in R_i} \exp(x_j^T \beta) \right) \right]$$

For computational efficiency, cox partial likelihood is usually minimized with mini-batch stochastic gradient descent (SGD), which is proved to be acceptable. Also, one of the disadvantage of traditional cox regression model is all patients shared the same baseline which does not depend on the time interval. To ease the limitation and integrate the time-related features, Kvamme et al introduce the time interval features into cox regression model [10]. Also, one model integrate the recurrent neural network (RNN) model to better represent the dependency between the sample and also used the censored patients for calculate the loss and back-propagation, which is more robust algorithms [11].

1.3.3. Pathology integrated with time-to-event analysis

Deep learning model that can handle the pathology images to predict the prognosis of patient get the interest because of the characteristics of dataset. If we predict accurately enough to define the new biomarker in a data-driven manner, it can bring the advantage for both patient and clinical pratitioners. Pioneering work combined the image features extracted using convolutional neural network (CNN) with

genomic features to predict the patient survival with cox partial likelihood function as the loss for the entire neural network model [12]. The proposed model can learn the visual patterns and molecular biomarkers associated with patient outcomes simultaneously. Also they provide insights into the prediction mechanisms of proposed method using heat map visualization to show that proposed method recognize important structures like microvascular proliferation that are related to prognosis and that are used by pathologists in grading. Abovementioned method used the pre-defined region of interest (ROI) to select the appropriate pathological patches from WSIs. Because the CNN has the limited receptive field with current graphical process units (GPUs), handling the entire WSI solely using the CNN is impossible. To tackle the above problems Stang et al used the pre-trained CNN to extract the features from several patches that randomly selected from whole slide image. And then, used the autoencoder to compress the features further with combined MLP and cox regression loss to predict the survival of patient [13].

All the abovementioned method showed great promises in terms of applying deep learning combined with cox regression to handle the pathological features. However, still the entire WSI cannot used for prognosis prediction and microenvironmental (context) features are not included in the analysis. There are still room for further improvement and in chapter 2, the remained technical challenges are discussed.

Chapter 2.

Current challenges for time-to-event prediction in pathology

Time-to-event prediction in pathology could bring new pathological biomarkers to help the patients and medical practitioners for better managing the disease. However, time-to-event prediction in pathology still have several challenges to achieve the advantages it can bring. In this chapter, difficulties to process the whole slide image as a semi-supervised learning because of patient-level label and large file size will be discussed. And difficulties to learn and interpret the context features which is important biomarkers in tumor microenvironment using graph neural network are introduced. Finally, the contribution of my works which is introduced in this dissertation in terms of tackling abovementioned challenges in detail.

2.1. Semi-supervised learning of whole slide image

2.1.1. Multiple instance learning

If the goal is to process a full slide image of each patient consisting of several small patches and classify patient-level features (staging, subtype) or predict the prognosis using only patient-level labels, then this case is interpreted in semi-supervised learning and multi-instance learning (MIL) practices. In the standard supervised learning case, one class label is assigned to each observation or "instance". However, in MIL, class labels are assigned to collections of observations or "instance bags". Therefore, compared to standard supervised learning, an additional challenge is to discover which instances of each bag affect the class label [14]. Likewise, in the medical image, effectively select the related instance (small-patches) from the whole slide image is challenging problem. Researchers try to pick the important small patches from the bags of instance using the sampled patch from unsupervised clustering method [15]. Recently, attention-based method, which is one-to-one measure of correlation with each combination of instances in bags and assign the appropriate weight to pick the important instances, can be integrated as the end-to-end manner with the neural-network and attention module naturally select the important instance in the bags of instances [16].

Lu et al used the pre-trained CNN to extract the compressed features from small patches that comes from WSI. And then, they apply the WSI-level attention module to classify the subtype of several cancers and showed promising results. Also, because attention score naturally represent the important rate of each patches they provide the attention score as the interpretation method.

Although we can handle the whole slide image in terms of MIL in end-to-end manner, it's hard to process the entire WSI with deep learning model because of limitation in resources of GPU. I will discuss more methods to handle the entire WSI in an end-to-end manner integrated with deep learning model.

2.1.2. Previous attempt to treat the whole slide image

If the user can define the region of interest (ROI) based on the pathologist label or another algorithm that can detect the ROI, I don't have to treat the WSI at once [17]. However, labeled dataset is both labor and cost intensive therefore it's unrealistic to secure the well labeled dataset for entire WSI dataset. In terms of above point, analyzing WSI with deep learning models presents another technical challenge. As mentioned above, the entire slide image can be up to 200,000 pixels in total. Therefore, gigapixel images are too large to fit on the GPU at once and usually, WSI is broken down into smaller patches for training deep learning models. The simplest way to process a WSI image is to randomly sample a small patch from the entire slide image [18]. However, the above approach does not cover the full diversity of full slide images, an important aspect of pathology. Image patches can be clustered and then sampled to capture a more varied view of the tissue shape. This can be done by individually clustering the patches in each image, one patch sampled from each cluster. The alternative to subsampling image patches is to use the entire WSI but in a compressed representation. A CNN can be trained on another dataset, which is pre-trained for large dataset such as ImageNet and used to encode small histopathology patches. Usually histopathology patches have $256 \times 256 \times 3 \approx 2 * 10^5$ dimensions and encoded patches have $1024 \approx 10^3$ which is two-fold decrease of entire dimensions [19]. I then need to predict the

class by aggregating the feature encodings of multiple patches in the WSI. These aggregation methods can be used to train an end-to-end CNN, RNN, or attention-based module using a subset of image patches, or used as a second-step model. Although the above methods suggest a promising way to treat WSI, it is a contextual feature that still has limitations in analyzing WSI like a pathologist. Describes the importance of contextual features and how to learn contextual features in a data-driven way.

2.2. Learning context features

2.2.1. Network representation of pathology image

Although promising, computational pathology remains limited in its acceptance of contextual features to analyze complex histopathological features in the tumor environment, such as pathologists. The conventional method used Convolutional Neural Networks (CNN) to extract features of small patch images (256x256 pixels) by segmenting the WSI. These methods naturally focused on local features such as morphological changes of tumor cells, the presence of various growth patterns of tumor cells, and structural features, which are important prognostic marker. [20], [21]. However, since these features can be explained differently depending on the surrounding circumstances, the pathologist examines different areas of the WSI at different magnifications and simultaneously considers contextual and local features to determine the patient's prognostic features.

For example, hemorrhage may be a clue of a poor prognosis if observed inside the tumor area, and may be a clue of a good prognosis if observed in normal cells. Immune cells may also have completely different roles depending on the tumor

microenvironment (TME), which is defined as the interaction of various cell types, such as tumor cells, lymphocytes, and stromal cells, present locally in the tumor [22]. Recent methods have attempted to solve the problem and hide contextual features by introducing a graph structure, which is an optimal data structure for analyzing contextual features by considering the interaction of different pathological features. To take full advantage of the rich contextual capabilities of WSI, I implemented classification, prognostic prediction, and more with graph neural networks (GNNs), a type of neural network that processes the dynamic structure of graphs and achieves state-of-the-art results in tasks such as tumor subtypes [23]–[26]. I will describe the more detailed explanation of GNN and advanced model of it to train the context features of pathology image.
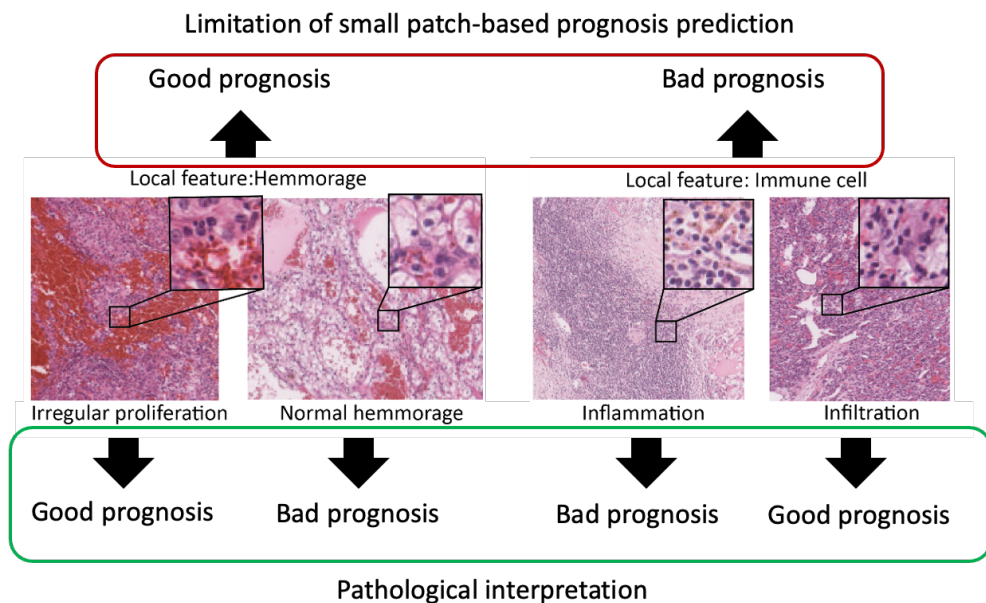
Figure 2.1 Advantage of network representation of pathology image

2.2.2. Graph neural network for context feature learning

Graph neural networks are deep learning-based methods that operate in the graph

13

domain. In recent years, the development of deep neural networks such as Multiple Layer Perceptron (MLP), Recurrent Neural Network (RNN), and Convolutional Neural Network (CNN) is rapidly increasing. In particular, CNN can extract multi-scale localized spatial features and construct them to construct highly expressive expressions. However, CNNs can only work with plain Euclidean data like images. In particular, it is difficult to define localized convolution filters and pooling operators in the graph region. Another motivation is learning graph representations, learning how to represent graph nodes, edges, or subgraphs as low-dimensional vectors like Deepwalk and Pagerank [27], [28]. Following the idea of representation learning and the success of word embedding, Deepwalk was considered the first graph embedding method based on representation learning.

Standard deep learning toolkits such as CNNs, RNNs, and MLPs include methods with various forms of relational inductive bias, but there is no standard for arbitrary relational structures that are graph structures. To solve the above problems, neural networks that use operators in graphs have increased dramatically in scope and popularity in recent years. Here, we discuss generalized messages conveying neural networks, a widely used framework for processing graph-structured data.
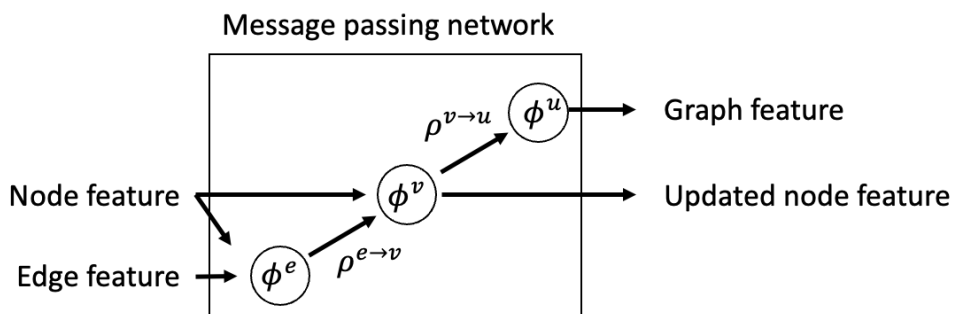
Figure 2.2 Principles of message passing neural network

Generally, message passing neural network consist of two aggregation function and

three update functions. Graph data consist of edge, node, and global-level features and the adjacency matrix. Adjacency matrix is defined the how to aggregate the node level and edge level features throughout the connectivity. Edge-level update function (ϕ^e) is applied for each edge and those messages (updated edge-feature) are passing through to the node that connected with each edge. And then, edge-level aggregation function ($\rho^{e \to v}$) is used to aggregate the edge-level feature to incorporate into the node-level feature. With that concatenated node-level and edge-level feature, node-level update function (ϕ^v) is applied for each node and derive the updated node-level feature. Finally, node-level features are aggregated, in other words readout to represent the entire global level graph feature through node-level aggregation function ($\rho^{v \to u}$). If we go further with the graph-level prediction, we performed the graph-level update function (ϕ^u) for graph-level decision [29]. For the aggregation function, we generally use the any pooling method such as max, min, average but we can also use the attention-mechanism or RNNs to secure more express power of model.

We can use the attention-mechanism to the aggregation function. Attention-mechanism is the special-case of the weighted sum where each weight is trained based on the correlation of each features. In the case of GNN, it calculate the attention-value (weight) based on the correlation with the connected neighborhood nodes after the edge-level and node-level feature update.

Figure 2.3 Schematic of attention mechanism of graph attention neural network

The basic backbone GNN model that used the attention mechanism is the graph attention network (GAT). At the GAT model, simplified attention-mechanism is utilized to secure both express power and computational cost. It conduct the self-attention mechanism within small subgraph (connected neighborhood nodes) and using the calculated attention value, it aggregates neighborhood node feature and updates target node feature. It can be applied to graph nodes having different degrees by specifying arbitrary weight to the neighbors. In the advanced form of GAT, we can use the edge features or also advanced form of self-attention. I will further discuss it in chapter 3.

2.2.3. Previous approach to apply graph neural network for pathology

Initially, graph neural network didn't apply to the whole slide image at once. Instead, Li et al construct the graph with randomly sampled patches from whole slide image [30]. This approach loose the spatial relationship between the small patches, which is the important context features within tumor microenvironment, thus can't extract meaningful context feature from tumor tissue. Recently, there are approaches that

based on graph neural network to tackle the whole slide image. First they transform pathological images into patch-based graph, where nodes are small patches and edges encode the relationships between the patches. Then, CNNs are used to extract features from these patches to generate a feature vector for the node embedding of the graph representation. Given the constructed graph, a graph deep learning model is used to conduct node or graph classification. Specifically, some methods handle the entire whole slide image at once for the graph deep learning model training.

Lu et al introduced a pipeline to construct a graph from the entire WSI using the nuclei level information [31]. They segment and classify the individual nuclei and extract feature from them. Then, agglomerative clustering is used to group spatially neighboring nuclei into clusters which results in reduced computational cost for downstream analysis. They evaluate the proposed framework for the breast cancer tissue and predict the status of human epidermal growth factor 2 (HER2) and progesterone receptor (PT) expression from WSIs of H&E stained tissue, which is supervised learning method.

Richard et al proposed semi-supervised based prognosis prediction method based on graph neural network [32]. Unlike Lu et al, they doesn't include the clustering method to construct the graph. Instead, they use the low-level magnificent (20x) and use the small-patch as the node in the graph. Because prognosis-related label (survival day) is patient-level label and node or patch-level label doesn't exist, Patch-GCN is the semi-supervised learning method. One interesting point in Patch-GCN is they incorporate global attention pooling method, therefore use the attention score as the interpretation method to find the pathological features related to prognosis.

17

2.3. Interpretability of graph neural network

2.3.1. Interpretation method for neural network

Neural networks have shown great promise in a number of tasks in the medical field. Interpretation of neural networks is important in many fields, but one area where interpretation is important is the medical field. In the medical field, decisions made by deep learning models usually relate to whether or not a patient's health condition is serious. Therefore, it is important to provide humans with interpretable evidence to support their decisions in a rational way. Before discussing interpretation methods for graph neural networks, I first briefly discussed interpretation methods for deep neural networks. Although there are several descriptive techniques, they can be grouped into three techniques: local surrogate, occlusion analysis, and gradient-based techniques.

Interpretable local surrogates aim to replace decision-making functions with self-explanatory ones, sometimes simpler models than target models for explanatory purposes. A popular method for this is the LIME algorithm [33]. The explanation can be achieved by first defining some local distribution around the data point x, learning the parameters v of the linear model that best match the function locally, and then extracting the local feature contributions. Occlusion analysis is a specific type of perturbation analysis that repeatedly tests the effect of occluded patches or individual features of an input image on the output of a neural network. A heatmap can be built from these scores highlighting the locations where functionality has decreased the most due to blocking. Shapley values can also be viewed in occlusion analysis. Instead of blocking features one at a time, a much broader set of blocking patterns is considered here. SHAP is a practical algorithm that approximates Shapley

values by sampling several occlusions according to the probability distribution used to compute the Shapley values and then fitting a linear surrogate model that correctly predicts the effects of these occlusions [34]. A unified gradient is described by incorporating a gradient along some trajectory in the input space that connects some root points to data points. In the case of graph neural networks, this method is of particular interest to the unified gradient method because it can easily be extended to any model.

2.3.2. Appropriate interpretation method for graph neural network

Interpreting graph neural networks is an important but difficult task. Graphs are not data like grids, unlike images or text. That is, there is no local information and each node has a different number of neighbors. Instead, graphs contain important topological information and are displayed as feature matrices and adjacency matrices. Also, graph data is less intuitive than images and text. For images and text, understanding the meaning of the input data is straightforward and straightforward. However, graphs can represent complex data, making it difficult for humans to understand the meaning of graphs. In addition, there are many unsolved mysteries among interdisciplinary fields such as chemistry and biology, and knowledge of the field is still lacking. Therefore, it is not easy to get a human comprehensible explanation of the graph model. The first approach is a perturbation-based method. Rex et al. proposed the pioneering work of graphing neural network interpretations called GNNExplainer [35]. The GNNexplainer takes a trained GNN and its predictions and returns an explanation in the form of a small subgraph of the input graph, along with a small subset of the node features that most influence the prediction. Given an input graph, I get a mask to represent the important input

19

features. Next, the generated mask is combined with the input graph to obtain a new graph containing the important input information. Finally, the new graph is fed to the trained GNN to evaluate the mask and update the mask generation algorithm. Benjamin et al compared several interpretation methods that actually help in choosing an appropriate interpretation method for graph neural networks [36]. Quantitatively evaluate multiple interpretation methods based on the four axes of accuracy, stability, fidelity, and consistency. Accuracy indicates how well the method matches the group-truth credit allocation. stability. Fidelity indicates that the performance of the method should reflect the performance of the original model. In conclusion, the use of CAM is recommended. However, CAM is not compatible with GNN architectures because it requires a global pooling layer as the last layer in the architecture. In these cases, IG is the right choice.

2.4. My works in this dissertation

Proposed graph neural network based whole slide image analysis technologies in this dissertation is a new approach that tackle the abovementioned problems and refers to tumor environment-associated context analysis using deep graph learning. The proposed method extracts the "contextual pathological features" of WSIs in a memory-efficient and semi-supervised manner and demonstrate it for the survival risk prediction problem of various cancer types.

The technologies in this dissertation address the following problems to analyze contextual pathological features which previous deep-learning models has difficulties.

(1) Several studies claimed that "contextual pathological features" can be learned, but none of them fully utilized the complex interaction of the tumor

environment that exists on the gigapixel size of the WSI. Previous methods used the features extracted with pre-trained convolutional neural networks or try to reflect the relationship through the recurrent neural network. However, previous approach lost the spatial information of cells through randomly selected regions of interest (ROIs) or lost the other pathological features by segmenting the cell nuclei, which are important contextual features.

(2) Other research used the pre-trained CNN feature and attention model to address the WSI at once and therefore did not lose any contextual features in the WSI. Although pioneering work, attention mechanisms have limitations in reflecting contextual features. The attention model did not reflect the structural context inside the tumor environment. The tumor environment is a complex graph of each cell type and pathological feature. Both features, such as which cells interact with other cells and where they interact, are important context features. However, the attention mechanisms are interconnected with each other, therefore mixing the whole graph structure. Therefore, we need to learn the appropriate contextual features that reflect the interaction of each cell in the tumor environment.

(3) Early attempts to introduce the graph representation of pathological images and implement the GNN to learn the appropriate context features are exist. But none of them learn the rich contextual features that reflect the complex interaction inside the WSI. To address the "contextual pathological features" of WSI, there are several technical bottlenecks. 1) The gigapixel WSI is too large to be trained at once in graphical processing units (GPUs). 2) Graph neural networks (GNNs) are promising candidates to learn "contextual

pathological features". However, as WSI has heterogenous pathological features, how to adapt the GNN appropriately to WSI is unknown. 3) It is hard to interpret the relationship among the pathological features that are learned through the GNN in the WSI.

The proposed TEA-graph utilized the WSI at once without losing contextual information on the WSI through a network compression technology named superpatch. Therefore, the TEA-graph only requires single GPUs to train the 1,000 WSIs in a day. I demonstrated the applying interpretable GNNs to WSIs. I approached the above difficulties through proposed unique solutions, which make it possible to mimic the "thinking" process of pathologists. 1) Memory-efficient network representation of the WSI that preserves the features in the original WSI. 2) Position-aware graph attention network (GAT) that reflects the heterogeneous tumor environment in the WSI. 3) Integrated gradients (IG) with attention score to simultaneously interpret the important pathological features and interaction of each pathological feature.

I demonstrated the proposed method for the survival and metastasis risk prediction problem of clear cell renal cell carcinoma (ccRCC), which has well-known "contextual pathological features" for prognosis prediction. I used 831 WSIs collected from Seoul National University Hospital (SNUH) and 502 WSIs from The Cancer Genome Atlas (TCGA) datasets for training and validation of the TEA-graph. I showed that the TEA-graph differentiated the heterogeneous tumor environment in the WSI and outperformed the conventional pathological grade. In addition, I suggest that "contextual pathological features" are related to the risk of survival and metastasis through interpretable visualization of WSIs. To interpret the WSI, I applied both the external interpretation metric IG and attention score as the

intrinsically learned interpretation metric.

Through this approach, I can measure the different importance of each contextual feature and introduce the interpretable pathological features with it. Overall, the TEA-graph is scalable to massive WSIs without manual annotation of experts and interpretable to researchers and clinicians. I expect TEA-graph to be easily accessible pathologists to analyze the pathological context.

In chapter 3, I will explain the overall concept of proposed method and showed the advanced graph attention network which is optimized structure for context features in tumor tissue. Lastly I will showed the performance in various tissue datasets which is evidence of proposed performance.

Chapter 3.

Context features with whole slide image

Incorporating context features within whole slide image to predict time-to-event is essential to reflect the tumor microenvironment appropriately. In this chapter, tumor environment-associated context analysis using deep graph learning (TEA-graph) is introduced which is a framework based on a memory-efficient graph representation of whole slide image and interpretable graph neural networks that considers the contextual features of the tumor environment in a semi-supervised manner. Performance of TEA-graph compared with conventional pathological marker and state-of-the-art deep learning model is presented to show the superiority of TEA-graph. Also using the interpretable method, novel pathological marker is introduced through TEA-graph.

3.1. Tumor environment-associated context learning using graph deep learning

3.1.1. Overall framework of proposed method
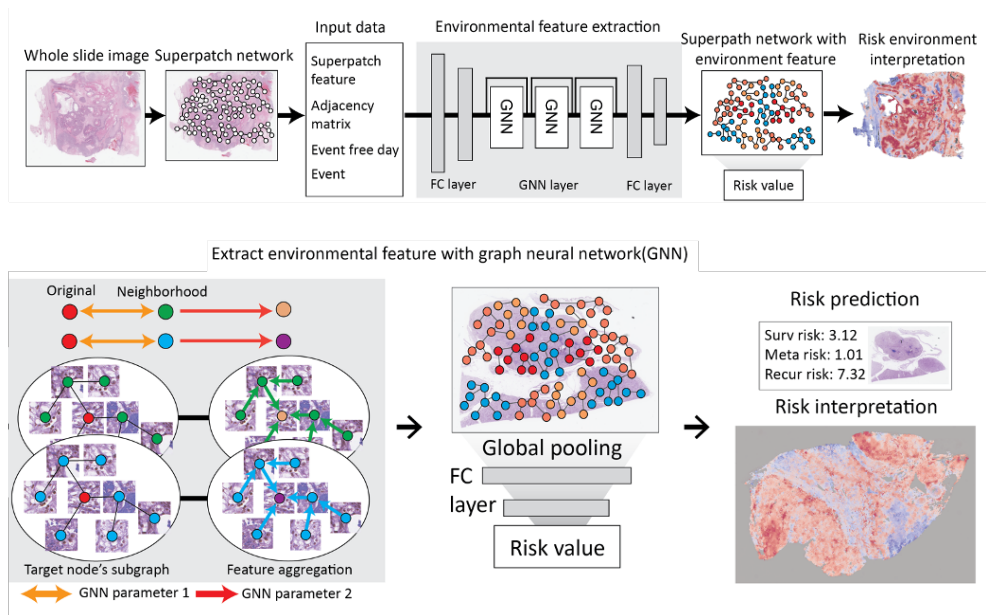
Figure 3.1 Proposed method to analyze the enviromental feature in whole slide image

I introduced contextual learning related to the tumor environment using graph deep learning (TEA-graph), a GNN-based method for semi-supervised analysis of the contextual histopathological features of gigapixel WSI. TEA-graph expresses WSI with superpatch, which is a memory-efficient WSI expression method that I propose. TEA-graph uses the full WSI to extract pathological context features through the GNN model while maintaining the spatial relationship of each local feature, and provides interpretability with intra-edge Attention Score and IG method.
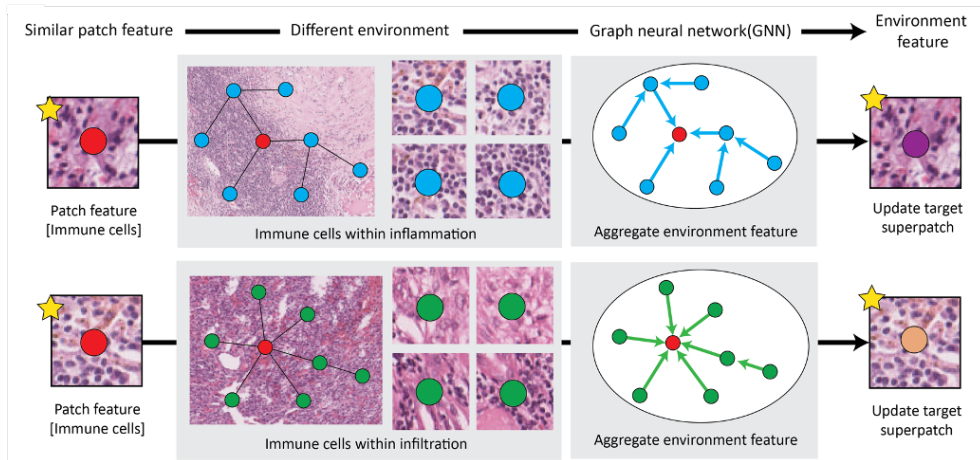
Figure 3.2 How graph neural network extract environmental feature from whole slide image

Although similar concepts of using graph neural networks for contextual learning and compressing whole slide images (WSIs) in various ways have been introduced, none of them introduces geometric features (positional features), which are important features that distinguish different tumor microenvironments.
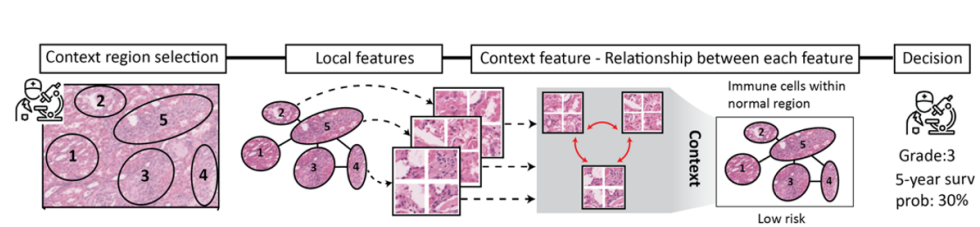
Figure 3.3 Schematic example of pathologist's decision making considering context features

In addition, I proposed an interpretation method of the context learning method, which is an essential aspect in the clinical application of deep learning. I demonstrate the use of the TEA-graph to predict the risk of an event (such as death or metastasis) and to analyze the pathological contextual features associated with the prognosis. TEA-Graph allows experts to train on the entire WSI without manually annotating the ROI and extract interpretable histopathological prognostic markers. I analyzed

26

risk-related characteristics and stratified the risk of patients more clearly than conventional histological grades.
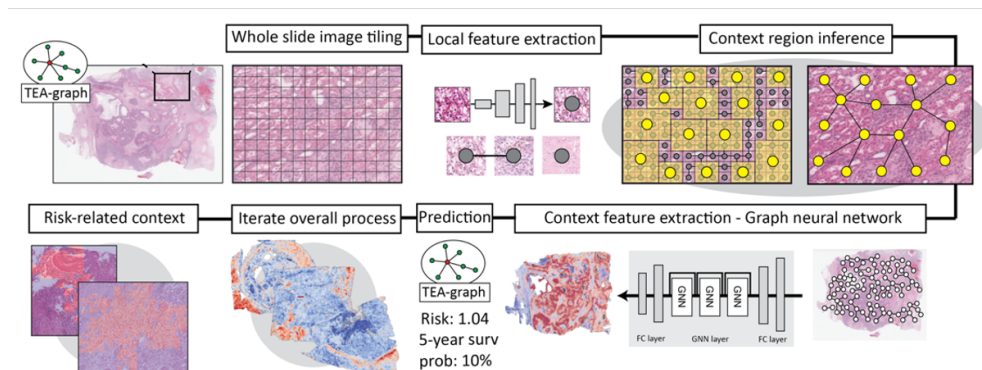
Figure 3.4 Proposed method mimic the pathologist's decision making method by graph neural network

I hope that the proposed TEA-graph method will be a useful tool for pathologists to find contextual pathological biomarkers. Therefore, I propose the analysis results of the prognosis-related pathological features extracted by TEA-graph. I showed the generality of the extracted pathologic features by finding similar pathologic features with high correlation with the pathological features extracted by TEA-graph in all patients. And patients with the pathological features suggested by the TEA-graph clearly showed adverse outcomes. In addition, I propose angiogenic pathological characteristics as an interesting biomarker for the adverse prognosis of renal cell carcinoma revealed by the proposed TEA-graph method. Overall, I propose a method to analyze TME for contextual features similar to that of a pathologist.

3.1.2. Super-patch network as whole slide image representation

One of the major bottlenecks for semi-supervised contextual feature learning in WSI is the limitation of the graphics processing unit (GPU) memory for processing gigapixel-sized WSIs. To address the memory issues associated with WSI while

accommodating contextual features, I use a super-patched approach to represent WSI as a graph structure. I apply the supernode method to WSI to compress gigapixel images and present them as memory-efficient graph structures. [37], [38]. Divide the WSI into smaller patches and extract the features of each patch using a pre-trained CNN model. A superpatch graph is constructed by repeatedly aggregating neighboring patches with similar characteristics in all WSIs. Therefore, the superpatch graph efficiently compressed the data size of the graph while preserving the spatial relationships between small patches.
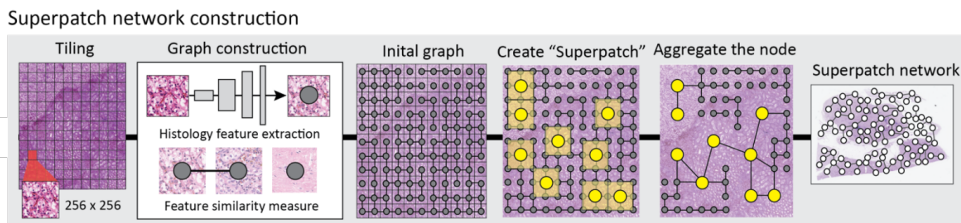
Figure 3.5 Superpatch network construction method

I confirmed that the superpatch wells represented nearby aggregated patches. I also compared the proposed superpatch method with several other superpatch construction methods. It was confirmed that the proposed method compresses the WSI to 10-40% of the original WSI while maintaining the information and features appropriately. Therefore, the superfetch method can be used to process the entire WSI in a memory-efficient manner with minimal loss of context information.

For superpatch generation, I first preprocess the WSI with an Otsu threshold to filter out artifacts in the WSI [39]. Only patches made up of 75% or more of Oats Mask are selected. I then divide the WSI into smaller patches of 256x256 pixels at 40x magnification [40]. Next, I define similar patches within the segmented smaller patches. First, I extracted the pathological features from the last MLP layer of the

ImageNet pre-trained EfficientNet. For WSI in ccRCC, I further optimized the pretrained EfficientNet for pathological applications through transfer learning of the entire network for cancer/normal classifiers without freezing the layers. For this purpose, 2,000 small patch images of 256 x 256 pixels were prepared for each tumor and normal class extracted from WSI, manually annotated by the pathologist. Transfer-learning is an optional step for better performance of TEA-graph. Comparable results can be obtained without transfer learning on TCGA datasets other than ccRCC types. To define similar patches, I consider the spatial distance (pixel-level L2-norm distance) and cosine similarity of features extracted from a pretrained EfficientNet that represents the similarity between images. To maximize compression of similar patches, patches within two patch distances are compared for spatial correlation (up to 24 patches in a 5x5 patch, 1 patch distance equals 256 pixels). Then, I measure the cosine similarity of the patch features extracted from the last MLP layer of the pretrained EfficientNet and define similar patches if the cosine similarity is lower than a defined threshold. After similar patches are defined within the WSI, the patches are sorted by the number of similar patches in each patch. It then starts with the patch with the most similar patch and proceeds in order to remove similar patches belonging to that patch. This is called a super patch. After superpatches are created, I treat each superpatch as a single node in the graph. If the length of the 5 patch distance is less than the length, only the spatial distance between the patches is taken into account to generate the edges between the nodes. The node function of a superpatch is obtained by averaging the functions of the patches incorporated into the superpatch. Additionally, each edge has an edge feature that contains the geometry of the superpatch via patient-specific normalized spatial distance and angle between the two patches.

29
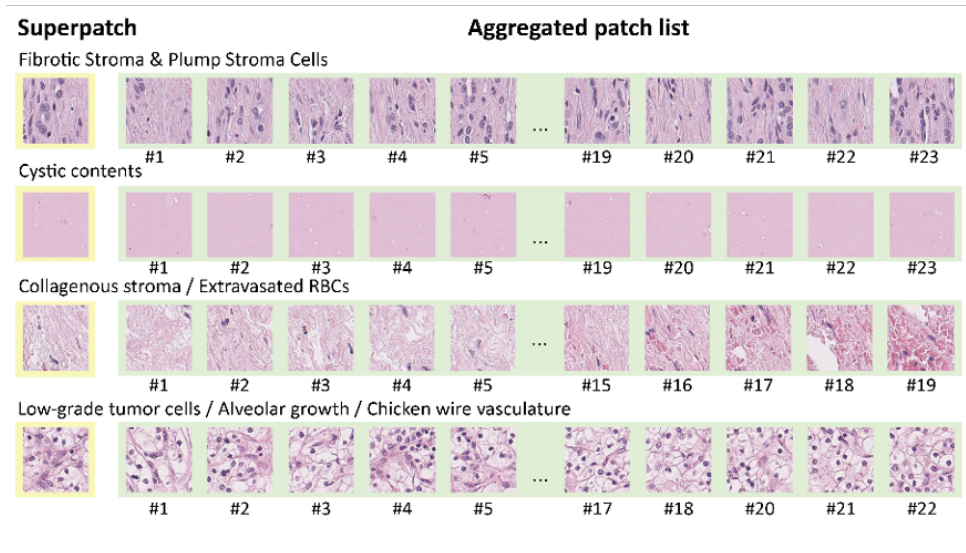
Figure 3.6 Example of representative superpatch and neighborhood aggreagted patches

To compare the compression ratio and the impact of threshold on information, I perform additional experiments comparing the compression ratio, label accuracy, and normalized mutual information (NMI) of the original WSI and compressed superpatch graphs at various thresholds. In addition, to prove that the proposed existing method is suitable for expressing WSI in a compressed graph structure, I compared the above metrics with those of the other four superpatch generation methods: original, frequency, length, and random methods.

Superpatch Number	Patch in Superpatch	Random	Length	Frequency
. . .	. . .	. . .	. . .	. . .
Patch 3451	{Patch 3340, Patch 3324, Patch 3401, Patch 3346, ...} N = 24	Random Selection N ~ U(0,24)	Random Selection N = 24	Random Selection N ~ P
Patch 3482	{Patch 3485, Patch 3478, Patch 3486, Patch 3501, ...} N = 24	Random Selection N ~ U(0,24)	Random Selection N = 24	Random Selection N ~ P
Patch 3543	{Patch 3556, Patch 3548, Patch 3532, Patch 3550, ...} N = 24	Random Selection N ~ U(0,24)	Random Selection N = 23	Random Selection N ~ P
. . .	. . .	. . .	. . .	. . .
Frequency	P~{24:0.03, 23:0.02, ...}	P ~ U(0,24)	P~{24:0.03, 23:0.02, ...}	P~{24:0.03, 23:0.02, ...}

Table 1 Three different method to make the superpatch by aggregating neighborhood small patches

The original method is proposed method, which generated super patches with small patches within a defined distance with cosine similarity values between super patch candidates higher than the threshold. The length model randomly samples small patches within a defined distance while maintaining the number of small patches grouped into super patches in the original method. The frequency model does not keep the number of small patches grouped into superpatches, but instead randomly selects the number of small patches while maintaining the overall frequency of a certain number of small patches aggregated by the original method. The random method randomly samples a random number of similar nodes and does not retain the conditions of the original method.

To quantitatively calculate the amount of compressed information, I calculated the similarity between the WSI and the compressed image with the NMI, compression ratio, and label accuracy. When the number of superpatches of the compressed image is N, I created an N×N matrix Q. Then, in the WSI, the superpatch is labeled according to the most common label of the patches included in the superpatch. After that, the (i,j)th component of the matrix indicates the number of superpatches where the WSI label of the superpatch is i and the label in the compressed image of the superpatch is j. The NMI between the WSI and the compressed superpatch graph is

$$\text{NMI} = \frac{-2 \sum_i \sum_j Q_{ij} \log\left(\frac{Q_{ij}N}{Q_{i.}Q_{.j}}\right)}{\sum_i Q_{i.}\log\left(\frac{Q_{i.}}{N}\right) + \sum_j Q_{.j}\log\left(\frac{Q_{.j}}{N}\right)}$$

where $N_{i.}$ and $N_{.j}$ are the marginal sums over the corresponding rows and columns and $N = \sum_i N_{i.} = \sum_j N_{.j} = \sum_{ij} N_{ij}$.

I also checked the compression rate by calculating the ratio between the number of patches in WSI and the compressed superpatch graph. To calculate the label accuracy, I checked whether the dominant label of the aggregated patches is same as the superpatch's label. Compared with random sampling, a threshold over 0.6 shows a significant difference. Therefore, for optimal performance, I set the cutoff as 0.75, and if I need to prioritize the compression ratio because of hardware limitations, I can choose a cutoff as 0.5, which performs better than the random selection method.

  Also, if multiple adjacent patches share the same number of similar patches, the superpatch can be randomly selected. However, even if the node selected as a superpatch changed, those alternatively selected superpatches were also positioned closely with the original selected superpatch, and those alternative superpatches also aggregated adjacent local patches, including the original patch selected as the

superpatch. Thus, whenever the selected superpatch node changed, the overall representation feature space was minimally affected. Therefore, in this revision round, I proved the rigidity of the superpatch network construction method. For that, I generated the randomized superpatch network on the same whole slide image.
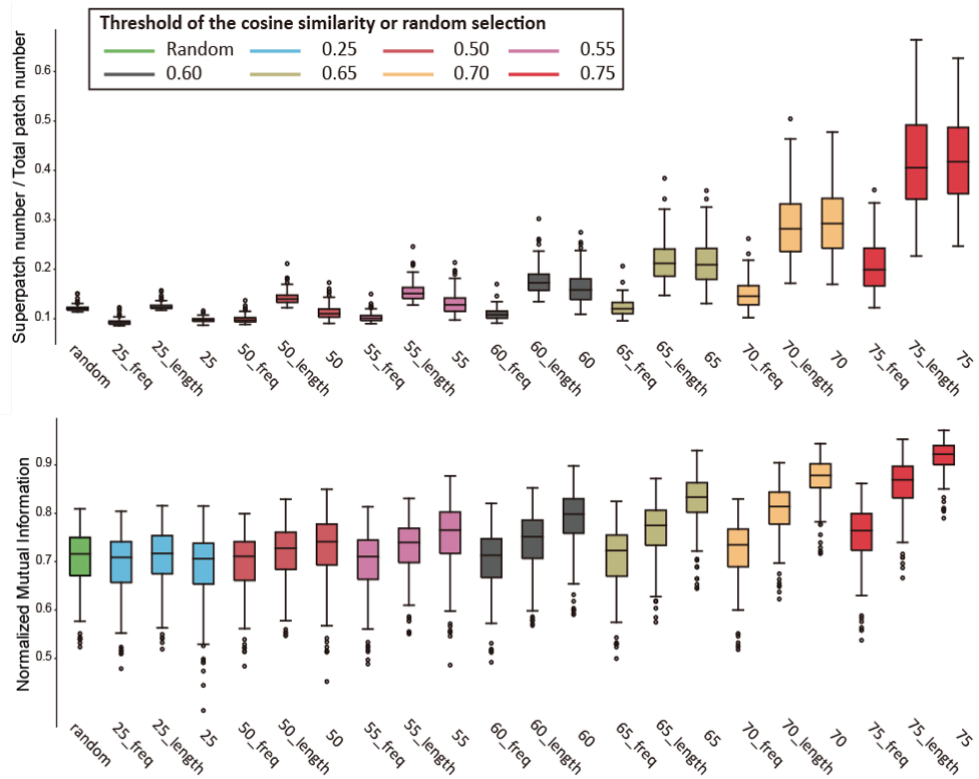
Figure 3.7 Validation of representative power of superpatch-network

The difference of the selected superpatch node was ~25%. However, as I expected, alternative superpatches represent features similar to those of the original superpatch network. The mean of the correlation between the alternative superpatch feature and original superpatch feature was very high (0.965), and the variance was low. I further checked that the rigidity of superpatch construction was different between the different pathological features. Because histopathological images per stage have different characteristics, I measured the correlation mean according to the TNM

stage. I again confirmed the high correlation between the alternative and original superpatches; thus, superpatch method was not affected by pathological characteristics and compressed the whole slide image appropriately.
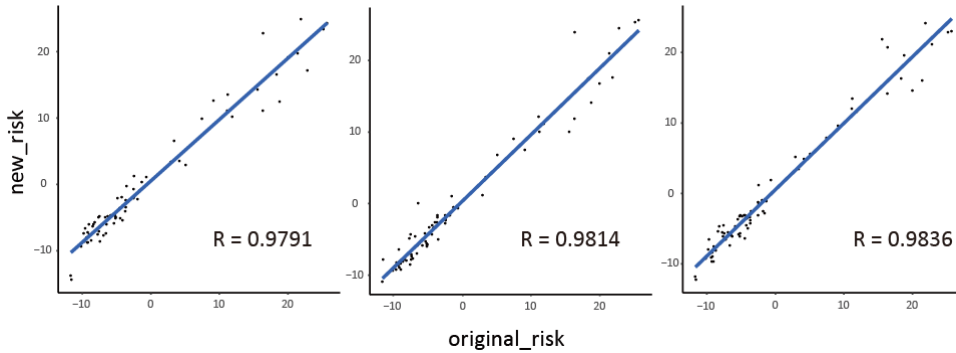
Figure 3.8 Correlation plot between the predicted risk using the original superpatch-graph and newly defined graph

To further confirm that alternately generated superpatches did not affect the risk prediction performance, I checked the predicted risk value's correlation between the three differently generated superpatch networks of 75 patients. The correlation between the original predicted risk score and the differently generated superpatch network was close to 0.98, which means that the alternately generated superpatch does not affect the risk prediction. Overall, I thoroughly tested the super-patch network as a method to represent the whole slide image efficiently within graph structure.

3.1.3. Position-aware graph neural network

I use a graph attention network (GAT) that utilizes attention scores within a GNN to learn contextual features in a heterogeneous tumor environment [41]. For example, when local pathological features are indicative of immune cells, these features may represent an entirely different histopathological context, such as inflammation or infiltration (i.e., tumor-infiltrating lymphocytes). Similar to how pathologists define

34

the prognostic value of these immune cells along with their surroundings, the GAT aggregates the neighbor features of the target superpatch with different weights for each neighbor. GAT updates the target superpatch function to include peripheral pathological features and to properly discriminate contextual features. Specifically, we modify the GAT to include the relative positions of each superpatch as edge features. The modified GAT utilizes the edge function to compute the value of interest between each node, thus better discriminating the relationship between each superpatch and representing the WSI better. Overall, TEA-graph takes a WSI as input and generates a compressed superpatch graph to handle multiple WSIs on the GPU. It then uses the contextual information to evaluate the risk value in the location-aware GAT.
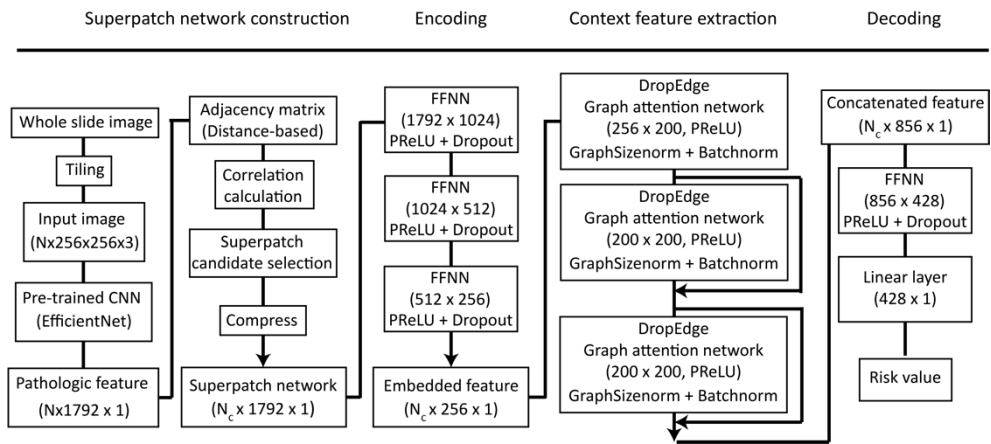
Figure 3.9 Overall pipeline of TEA-graph with input features and feature dimensions

The overall structure of the TEA-graph is as follows. To process WSI with end-to-end learning, I reduce the feature vector dimension to 200 using a three-layer multilayer perceptron (MLP) for preprocessing (800, 400, and 200 dimensions). I use GNN, an optimal neural network that processes graph structures for histopathological context learning. To ensure interpretability and maximize the

learning ability of GNNs, I choose GAT as backbone model. The model learns the attention score for each edge, so it differentiates the weights of each node when aggregating neighboring features. These attention scores are also used to interpret the significance of each connection between superpatches. Specifically, I modified the attention scoring method to include the geometrical features of edge features and developed a three-layer GAT with two attention heads (100 dimensions per head) in each layer as a basic structure with sufficient receptive fields to capture. Use. Environmental characteristics of GNNs. To represent geometrical features as learnable parameters, I quantized normalized distances and angles to 0-10 numbers and generated lookup tables with learnable features for distances and angles. In addition, after using LayerNorm, I activate for each layer to prevent overfitting, and use a parametric rectified linear unit (PReLU) as an activation layer and residual connections between each layer to improve learning ability and stability [42]–[44]. To calculate the risk score, I concatenate the output of each GAT layer, add two-layer MLPs (800 and 400 dimensions) as post-processing, and add a final fully connected layer that produces a single risk score as the output of the WSI. I use PyTorch and PyTorch Geometric for MLP and GAT implementation[45].

$$\boldsymbol{\alpha}_{ij} = Attention\ value, a_s, a_r, a_a, a_d\ = Attention\ weight$$

$$W_s h_i, W_r h_j, W_a a_{ij}, W_d d_{ij} = Transformed\ node\ feature$$
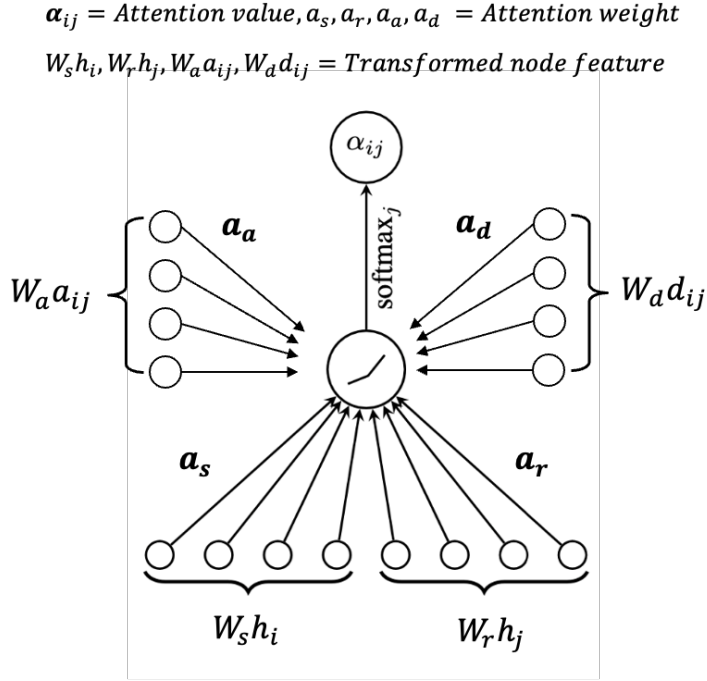
Figure 3.10 Proposed position-aware graph attention network

I suggest the detailed explanation of proposed position aware graph attention network. The input $\boldsymbol{h}$ to layer is a set of histopathological features h_i extracted from the last layer of the preprocessing MLP, where N is the number of superpatches (nodes), and F_n is the number of features in the last layer of the preprocessing MLP.

$$\boldsymbol{h} = \{h_1, h_2, \dots, h_N\}, h_i \in \mathbb{R}^{F_n}$$

I also embed the distance $\boldsymbol{d}$ and angle $\boldsymbol{a}$ as the edge features of the superpatch graph. I embed the distance and angle by the learnable parameters, where F_e is the number of learnable parameters for embedding the distance and angle features.

$$\boldsymbol{d} = \{d_{11}, d_{12}, \dots, d_{NN-1}, d_{NN}\}, d_{ij} \in \mathbb{R}^{F_e}$$

$$\boldsymbol{a} = \{a_{11}, a_{12}, \dots, a_{NN-1}, a_{NN}\}, a_{ij} \in \mathbb{R}^{F_e}$$

To reflect the heterogeneous surrounding environment features in the context features, I calculate the self-attention score α_{ij} between adjacent neighborhood

superpatches. Additionally, the spatial location of each histopathological feature is important to consider. Thus, I include the distance and angle feature to calculate the attention score so that positional information is passed between each superpatch. W_s, W_r, W_d, W_a are the initial linear transformations of the source node, receiver node, and distance and angle features shared in the layer, respectively. a_s, a_r, a_d, a_a represent a single-layer feedforward neural network for the attention coefficients for the source node, receiver node, distance, and angle features. $\mathcal{N}_i$ is the neighborhood superpatches of superpatch i in the graph.

$$\alpha_{ij} = \frac{\exp\left(LeakyRELU(a_s^T W_s h_i + a_r^T W_r h_j + a_d^T W_d d_{ij} + a_a^T W_a a_{ij})\right)}{\sum_{k \in \mathcal{N}_i} \exp\left(LeakyRELU(a_s^T W_s h_i + a_r^T W_r h_k + a_d^T W_d d_{ik} + a_a^T W_a a_{ik})\right)}$$

Once obtained, the normalized attention coefficients are used to compute a linear combination of the corresponding features. σ is the nonlinearity, and I used PReLU for sufficient expression. h_i' represents the final contextual features.

$$h_i' = \sigma\left(\sum_{j \in \mathcal{N}_i} (\alpha_{ij} W h_j)\right)$$

I use the Cox regression loss as objective to adapt the survival prediction into a GAT. The last fully-connected layer produces the predicted risk $R = \beta^T X$ associated with the input graph. $\beta \in \mathbb{R}$ is the last layer weights, and $X \in \mathbb{R}$ is the inputs to the last layer. These risks are input to the negative partial log-likelihood of Cox proportional hazards regression, which is the loss function for the model. U is the entire list of patients, and ϕ_i is the list of patients who have shorter survival times than patient i.

$$L(\beta, X) = -\sum_{i \in U} \left(\beta^T X_i - \log \sum_{j \in \Phi_i} e^{\beta^T X_j}\right)$$

The model updates the parameters to minimize the loss through backpropagation using the Adam optimizer. I apply the default parameters for the Adam optimizer

with a weight decay factor = 0.0005. Additionally, to allow the model to converge faster, I implement a learning rate scheduler that decreases the learning rate by a factor of 0.95 every ten steps.

3.1.4. Interpretable pathological context features

Additionally, for a thorough explanation of histopathological contextual analysis, I utilize an interpretive method called the IG method to highlight the risk values of each patch [46]–[48]. The IG value represents the positive or negative impact of a particular feature on the model. Interpret the risk value of each context feature by calculating the IG value for each superpatch, which is a node in the graph. Therefore, if a well-trained TEA graph accurately predicts risk, then the IG value can be correlated with the risk value of context features included in the trained superpatch. TEA-graph also provides sub-graph-level interpretability through connected graphs of groups of similar IG value nodes. Consequently, the TEA-graph suggests the prognostic relevance of the tumor environment represented by the connected graph. The IG method is used to evaluate the influence of input nodes in a graph. In case, each node has a node feature V of the pre-trained CNN model. For IG analysis, I need to define a baseline and interpolate the node features from the baseline feature $\tilde{V}$ to the original input node feature V. Calculate the IG using the IntegratedGradients function in the captum(0.4.0) module. The function uses a node function that is zero for the baseline function.

The IG value is calculated as follows:

$$\text{IG}(V^i) = \sum_{k=1}^{l}(V_k^i - \tilde{V}_k^i) \times \sum_{p=1}^{m} \frac{\partial GNN[\tilde{V} + \frac{p}{m} \times (V - \tilde{V})]}{\partial V_k^i} \times \frac{1}{m}$$

V_k^i represents the i-th node, k represents the k-th node feature, and m represents the

interpolation step. After calculating the IG of each node for comparison, I normalize all IG values across the entire WSI using min-max normalization. A node with a large IG value means that the node has a large influence on the prediction output (risk). The sign of the IG value indicates the direction of influence, and if the sign of the IG value is negative, it means that the direction in which the node affects the risk is the direction of decreasing risk. For validation, I collect IG values from the entire data set and group the top 10% (high), middle 10% (medium), and bottom 10% (low) of IG values. Use this group to characterize the histopathological features of each group. Use the median number of patches in each patient as a threshold to divide the patients into two groups using the IG group. After dividing the groups, a Kaplan-Meier survival analysis is performed to determine the prognostic characteristics of each IG group.

  Also, to check the correlation between the attention value and interpretability, I extracted high (>0.9) and low attention (<0.1) superpatch pairs and measured the correlation between features of each node taken from a pre-trained CNN. I calculated the fraction of high-correlation (>0.8) and low-correlation (<0.2) pairs of superpatches across high-interest and low-interest pairs. In addition, I extracted high attention and low attention superpatch pairs for each of the previously defined high, mid, and low IG groups. At this time, I calculated the median feature correlation values within the overall high and low attention pairs for each IG group.

3.2. Performance of TEA-graph

3.2.1. Comparison with standard pathology score

I used 5-fold cross-validation to evaluate the patient-level risk predictive

performance of the TEA-graph for three different events (survival, progression, and metastasis) in patients with ccRCC. I randomly split the WSI dataset of 831 ccRCC patients from Seoul National University Hospital into a training set (80%), validation set (10%), and test set (10%). The model was trained to predict patient risk values with all training set WSIs and evaluated with test set WSIs. The concordance index (C-index, method) of various histopathological data (WHO/ISUP grade, TNM staging and other metadata described in Methods) provided by the pathologist was compared with the predicted risk values obtained from the TEA-graph. TEA-graph showed limited performance compared to the TNM stage, which contains information that cannot be inferred from tissue images, such as tumor size, but TEA-graph outperformed the WHO/ISUP rating for all three events, with predictive accuracy of 88%. In addition, when combined with histopathological data provided by a pathologist, the TEA-graph predicted risk value performed better than all other histopathological data, confirming the usefulness of the TEA-graph. enemy data. In addition, the TEA graph predicted risk values achieved the highest risk ratios, showing that the predicted risk scores better reflect each patient's probability of occurrence of an event compared to data provided by other pathologists for all events. We also evaluated the performance of the TEA-graph for patient risk stratification by quantifying the predicted risk values into four separate groups to generate predictive histopathological grades. Predicted histopathological grades showed better stratification of patient risk than WHO/ISUP grades, particularly early-stage

patients for both survival and metastatic events.
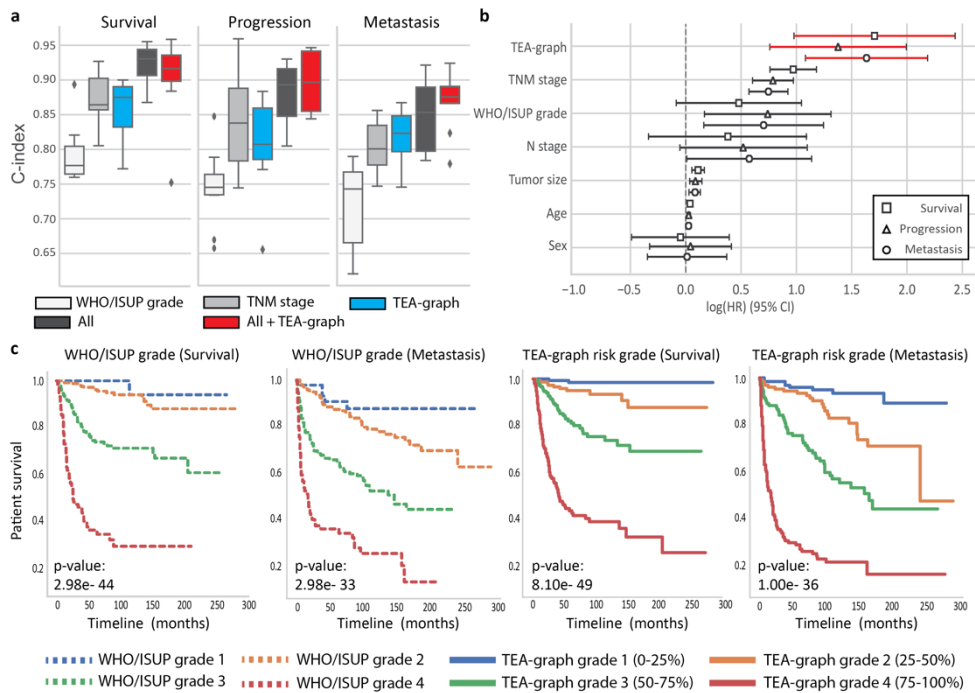
Figure 3.11 Comparison of the performance between the TEA-graph and other prognostic biomarkers. **a,** C-index of various prognostic biomarkers with TEA-graph-predicted risk value for three different events (survival, progression, and metastasis). All includes age, sex, WHO/ISUP grade, and TNM stage. Interquartile range (IQR) of box plot is between $Q1 – Q3$ and centre line indicates median value. Maxima is $Q3 + 1.5*IQR$ and minima is $Q1 – 1.5*IQR$. **b,** Log hazard ratio of several biomarkers with TEA-graph-predicted risk value in relation to survival, progression, and metastasis events. TNM stage and WHO/ISUP grade are the conventional prognostic markers that have been defined previously. All hazard ratio values were measured with SNUH dataset (n=831). Confidence interval was calculated by sandwich variance estimator. **c,** Kaplan–Meier survival analysis of survival and metastasis events using WHO/ISUP grade (left). Kaplan–Meier survival analysis of survival and metastasis events using the TEA-graph-predicted risk value (right). The TEA-graph grade was calculated by dividing the TEA-graph-predicted risk value into four groups using quartiles of predicted risk value. P-values were calculated through two-sided log-rank test.

3.2.2. Comparison with state-of-the-art model

I validated TEA graphs with external data sets and compared their performance with other multi-instance or contextual feature learning models [30], [49]–[51]. I trained and evaluated on WSI datasets of patients with kidney (KIRC), breast (BRCA), lung (LUAD and NLST), and uterus (UCEC), and found usefulness of TEA graph predictive risk scores compared to conventional risk scores.

Models	CCRCC	TCGA-KIRC	TCGA-BRCA	TCGA-LUAD	TCGA-UCEC	NLST	Overall
MIL	0.788 ± 0.032	0.617 ± 0.023	0.618 ± 0.027	0.642 ± 0.054	0.689 ± 0.065	0.701 ± 0.032	0.676
Attention MIL	0.804 ± 0.038	0.631 ± 0.021	0.739 ± 0.052	0.620 ± 0.116	$\mathbf{0.727 \pm 0.056}$	0.702 ± 0.049	0.704
DeepGraphConv	0.800 ± 0.021	0.657 ± 0.079	0.680 ± 0.013	0.687 ± 0.040	0.680 ± 0.074	0.683 ± 0.049	0.698
Patch-GCN (Originals)	N/A	N/A	0.58 ± 0.025	0.585 ± 0.012	0.629 ± 0.056	N/A	0.598
Patch-GCN (Modified)	0.775 ± 0.006	0.674 ± 0.044	0.704 ± 0.036	0.667 ± 0.053	0.667 ± 0.043	0.685 ± 0.032	0.695
TEA-Graph (GIN)	0.810 ± 0.014	0.708 ± 0.036	0.727 ± 0.043	0.677 ± 0.051	0.623 ± 0.068	0.704 ± 0.035	0.706
TEA-Graph (Ours)	$\mathbf{0.848 \pm 0.022}$	$\mathbf{0.710 \pm 0.013}$	$\mathbf{0.750 \pm 0.021}$	$\mathbf{0.709 \pm 0.016}$	0.687 ± 0.017	$\mathbf{0.731 \pm 0.009}$	**0.739**

Table 2 Performance comparison with other contextual and multi-instance learning (MIL) models

For each prognostic task (survival, progression, metastasis), calculate the calculated risk using multiple biomarkers (age, sex, WHO/ISUP grade, TNM stage) and concordance index (C-index) and TEA-graph. The C-index measures whether the sample is correctly aligned. Here, if one sample is predicted to have a higher risk than the other, then that sample should have a shorter survival time than the other. For example, two samples a and b match if the estimated risk and survival times t_a and t_b of each sample r_a and r_b satisfy ($r\_a>r\_b$ AND $t\_a<t\_b$). Otherwise, the C-index measures the ratio of concordance between two sample pairs to comparable sample pairs. Thus, a C-index of 1.0 means that all predicted risk values are correctly aligned with the survival time of each sample, and a C-index of 0.5 represents a completely random condition. Cox regression analysis was also performed to

calculate the risk ratio and calculated risk for each biomarker (age, sex, WHO/ISUP grade, TNM stage, lymph node metastasis, tumor size). A Kaplan-Meier survival analysis was performed using conventional histological grades and calculated risks. Quantize the patients into four groups using the quartiles of the calculated risk to combine the rating with the calculated risk. We selected the Lifeline (0.26.0) package in Python for survival analysis [52].

Dataset	Organ	Source	Event type	Total patients	Event number	Stage
SNUH_CCRCC	Kidney	SNUH	Death	831	138(17%)	1: 565(68%) 2: 84(10%) 3: 98(12%) 4: 84(10%)
			Progression		230(28%)	
			Metastasis		163(20%)	
TCGA_CCRCC	Kidney	TCGA	Death	502	167(33%)	1: 248(49%) 2: 52(10%) 3: 121(24%) 4: 79(16%) Unknown: 2(~1%)
TCGA_BRCA	Breast	TCGA	Death	1,098	152(14%)	1: 183(17%) 2: 621(57%) 3: 249(23%) 4: 20(1%) Unknown: 25(2%)
TCGA_LUAD	Lung	TCGA	Death	522	188(36%)	1: 279(53%) 2: 124(24%) 3: 85(16%) 4: 26(5%) Unknown: 8(2%)
TCGA_UCEC	Uterine	TCGA	Death	548	91(17%)	1: 342(64%) 2: 52(9%) 3: 124(23%) 4: 30(5%)
NLST	Lung	NLST	Death	449	159(35%)	1: 288(64%) 2: 64(14%) 3: 68(15%) 4: 25(6%) Unknown: 4(1%)
Total		-	-	4,947	-	-

Table 3 Entire tumor datasets used for training and testing

TEA-graph outperforms contextual feature learning models that do not utilize relative positions as edge features. We found that TEA-graphs in the KIRC dataset showed better prognostic performance than cancer staging and other histopathological data. The results were consistent with the SNUH data set.
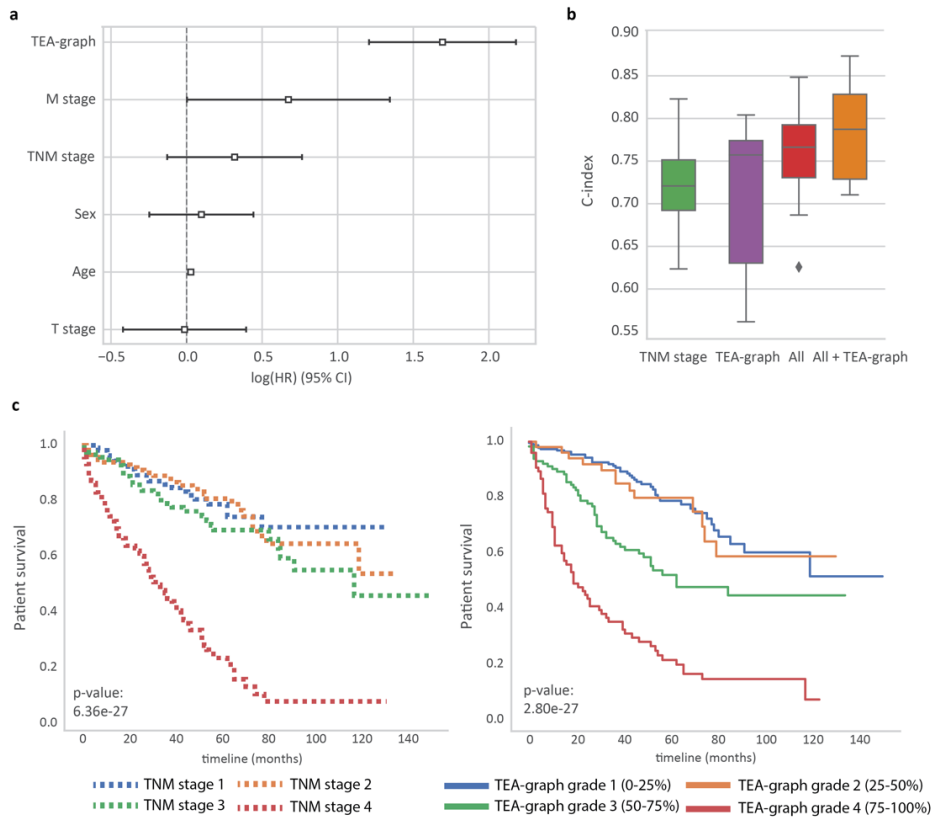
Figure 3.12 Validation of the TEA-graph on the external TCGA dataset. **a,** Log hazard ratio of several biomarkers with the TEA-graph-predicted risk value in relation to survival. All hazard ratio values were measured with TCGA dataset (n=502). Confidence interval was calculated by sandwich estimator. **b,** Concordance index (C-index) of various prognostic biomarkers with the TEA-graph-predicted risk in relation to survival. All includes age, sex, and TNM stage. IQR of box plot is between Q1 – Q3 and centre line indicates median value. Maxima is Q3 + 1.5*IQR and minima is Q1 – 1.5*IQR. **c,** Kaplan–Meier survival analysis of survival using the TEA-graph predicted-risk value (right) and the original stage (left). P-values were calculated through two-sided log-rank test.

3.3. Interpretation of TEA-graph related to pathology

3.3.1. Integrated gradients value as the valid tool for interpretation

Using the TEA-graph, I evaluated whether the calculated IG values adequately describe the risk-related characteristics. Superpatches with IG values of top 10%,

middle 10%, and bottom 10% were grouped into high IG group, medium IG group, and low IG group, respectively. In terms of prognosis, the low IG group and the high IG group have a good prognosis and a poor prognosis, respectively. The mid IG group represents a prognostic or prognostic-independent function between the low and high IG groups. First, it was confirmed whether the TEA-graph reflects local features related to prognosis by checking the superpatches of the high, medium, and low IG groups. I observed correlations between each IG group and WHO/ISUP ratings, indicating well-known prognostic features.

Figure 3.13 Risk-related histopathological characteristics predicted by TEA-graph. **a,** Histogram plot of the overall IG value that is correlated with predicted risk. **b,** Histopathological characteristics of small patches, which TEA-graph predicts as low risk (low IG), no risk (mid IG), and high risk (high IG). Scale bar, 30μm. **c,**

Kaplan–Meier survival analysis of each IG group (high, mid, and low) by dividing the patient into high-count/low-count groups for each IG group. P-values were calculated through two-sided log-rank test.
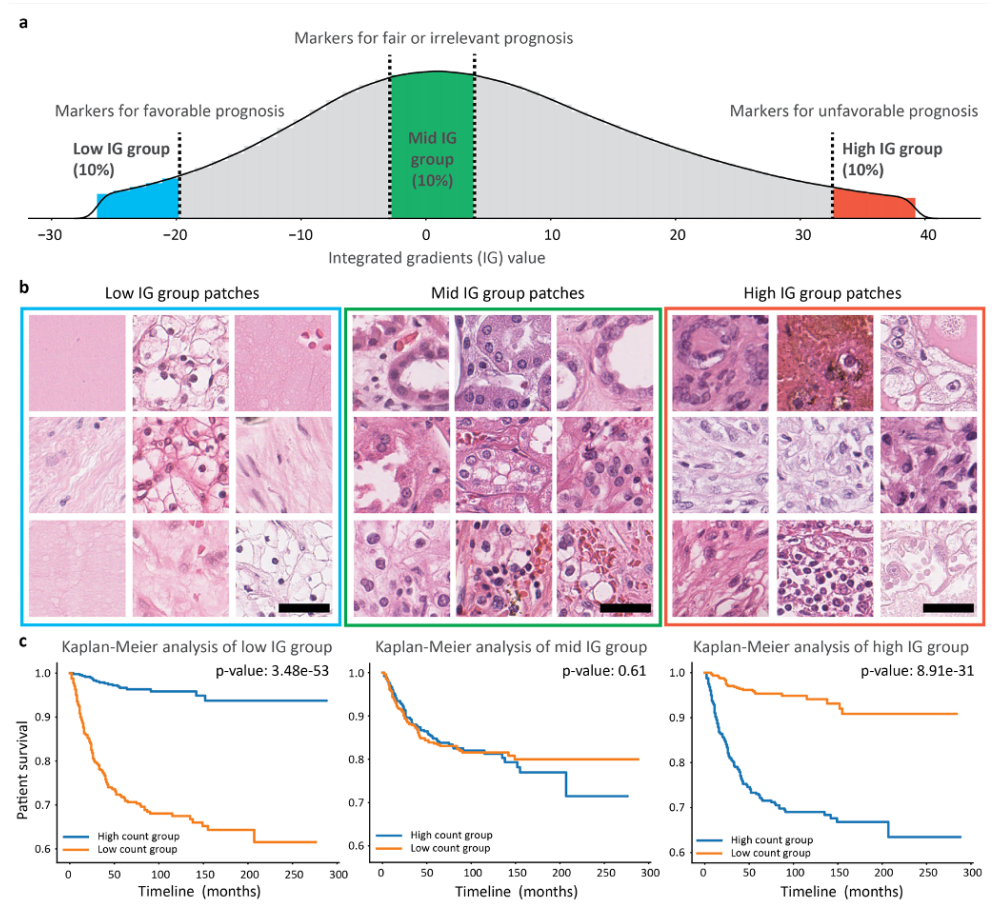
For example, I identified small, transparent tumor cells in the low IG group and polymorphic tumor cells or tumor cells with rhabdomayosarcomatous differentiation in the high IG group. In the mid IG group, I observed tumor cells with a clear or eosinophilic cytoplasm, and normal areas such as glomeruli or glomerular tubules. Second, it was checked whether the TEA-graph reflects the contextual features by considering the interaction with the surrounding local features. I extracted 3-hop subgraphs of each superpatch from the high-, mid-, and low IG groups. The three-hop is a subgraph boundary that the proposed GNN model can learn to describe contextual features. I observed cystic changes and stromal change in the low IG group, which are well-known contextual features for a favorable prognosis. In the mid IG group, I observed alveolar or papillary growths and stromal hemorrhage in the tumor and around normal component such as tubules and stroma, which are normal architectural contextual features. On the other hand, I identified hemorrhage, necrosis and lymphocyte infiltration as high-risk contextual features in the high IG group.

In order to prove that each superpatch IG value appropriately and quantitatively reflects the prognostic factors, the correlation between the average of the IG values and the predicted risk of the TEA graph was measured. IG values clearly correlated with predicted risk values and, as expected, each group of IG values was enriched in the sorted risk groups. Additionally, for each patient, I counted the number of superpatches in 3 different IG groups and stratified these patients into two groups: high-count patients and low-count patients in 3 different IG groups. I clearly observed that patients had high counts in the high IG group and vice versa, low

survival in the low IG group. In addition, I subdivided the quantized IG groups by 10% of the total IG values and then performed Kaplan-Meier analyzes of the high- and low-count groups in the subdivided IG groups according to the prognostic characteristics of the matched risk groups.
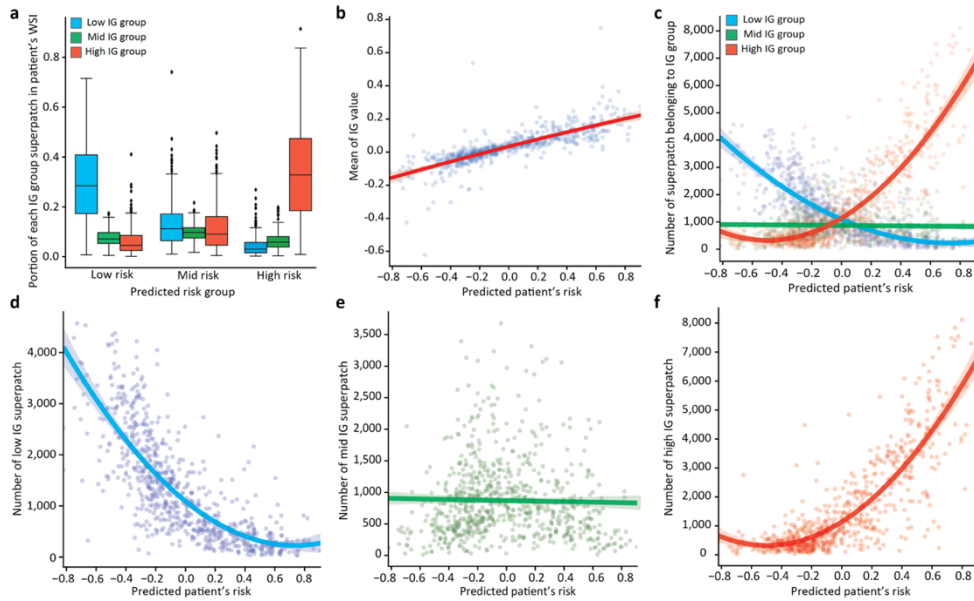
Figure 3.14 Correlation measurement between the risk and IG value. **a,** Numbers of patches in the low, mid, and high IG groups for each risk group. IQR of box plot is between Q1 – Q3 and centre line indicates median value. Maxima is Q3 + 1.5*IQR and minima is Q1 – 1.5*IQR. **b,** Scatter plot between the risk and IG values. **c,** Merged scatter plot between the risk values and numbers of patches in each IG group. **d-f,** Scatter plots between the risk values and numbers of patches for each IG group.

3.3.2. Validation of interpretability in two different tumor tissue

It is important to adequately predict risk values to reflect the heterogeneous tumor environment. Therefore, I examined whether the TEA-graph distinguishes two different contexts and reflects the risk prediction context through attention scores and IG values between each superpatch. I demonstrated that TEA-graphs discriminate histopathological features, with similar local features but different

surrounding environments using attention mechanisms. I observed that the TEA-graph could identify the various roles of immune cells according to the context of the tumor environment and adequately reflect it in risk prediction.
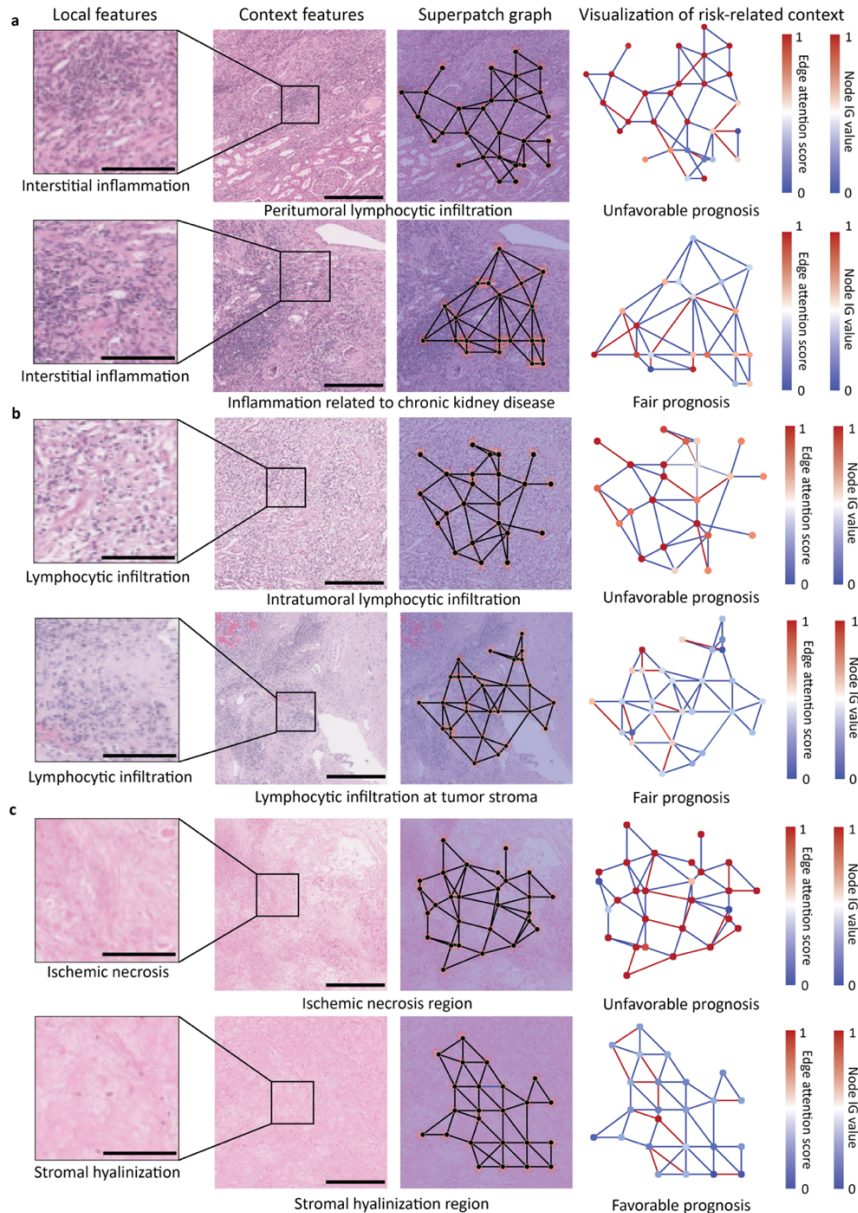
Figure 3.15 Heterogeneous context feature extracted by TEA-graph. First column shows patches that have similar local features and second column shows different pathological context. The color of each node indicates the node IG value and the color of each edge indicates the attention score. High attention score between two

nodes points out connected nodes affect each other significantly **a,** Peritumoral lymphocytic infiltration and compressed tubules and glomeruli (top). Interstitial inflammation and fibrosis related to chronic kidney disease (bottom). Scale bar, 100μm (left), 400μm (right). **b,** Intratumoral lymphocytic infiltration (top). Lymphocytic infiltration at fibrotic tumor stroma (bottom). Scale bar, 100μm (left), 400μm (right). **c,** Ischemic tumor necrosis (top). Stromal hyalinization (bottom). Scale bar, 100μm(left), 400μm(right).

For example, although interstitial inflammation was observed as a local feature, TEA-graphs identified other prognostic features between peritumoral lymphocyte infiltration and inflammation associated with chronic kidney disease. Furthermore, I observed that the TEA-graph differentiated prognostic effects when lymphocytes showed infiltration into tumor cells or were predominantly present in the tumor stroma with little association with tumor cells. In addition, TEA-graph showed a difference between ischemic tumor necrosis and stroma vitrification with similar local features, but with different prognostic effects.

Similarly, the TEA-graph imposed a different prognostic effect for each bleeding feature. I also investigated the role of attention values by observing feature correlations between two connected patches with low (< 0.1) and high (> 0.9) attention values. Interestingly, I found that the TEA-graph favored paying high attention between two patches with different (low correlation) features. This trend is more pronounced in the low IG group and the high IG group than in the mid IG group. In other words, the TEA-graph aggregates heterogeneous characteristics relevant to a patient's prognosis to generate contextual features. Overall, I demonstrated that TEA-graphs well reflect the contextual features of the tumor environment through IG values and attention scores, even in difficult cases.
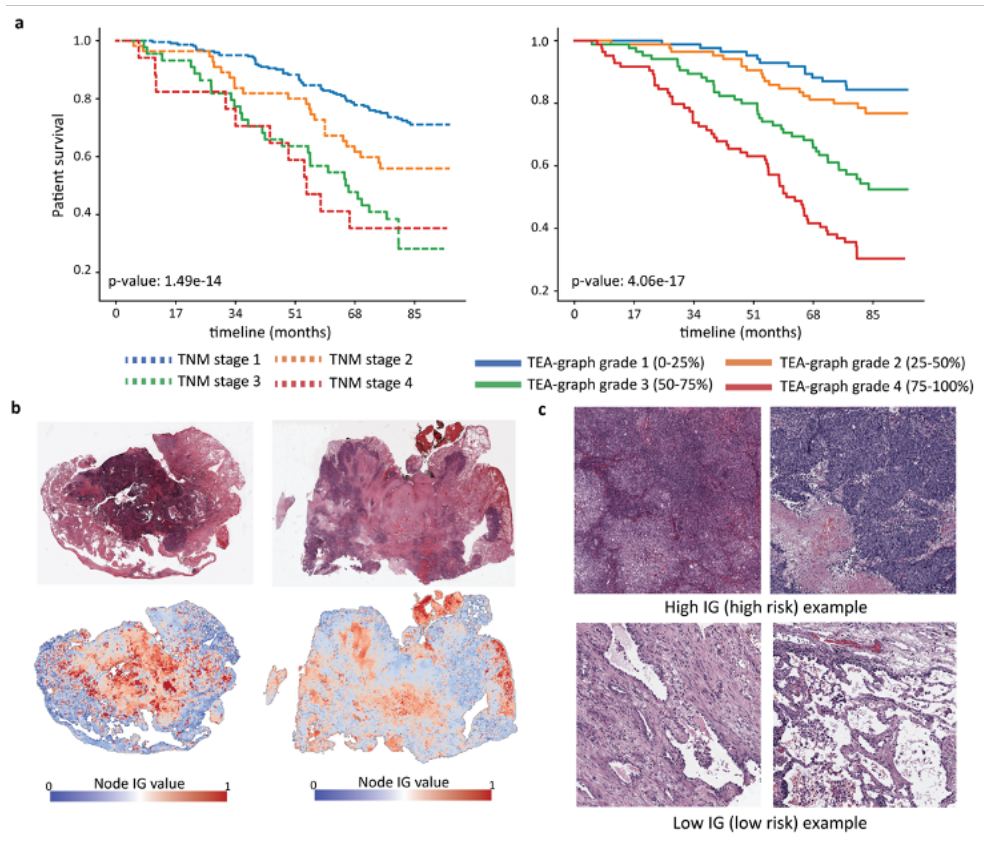
Figure 3.16 Validation of the TEA-graph on the external NLST dataset. **a,** Kaplan-Meier survival analysis of survival using the TEA-graph predicted-risk value (right) and the original stage (left). P-values were calculated through two-sided log-rank test. **b,** Predicted risk heatmap of NLST patients. **c,** Risk-related contextual features predicted by the TEA-graph.

I further strengthened the validity of IG values as risk interpretation values by examining external data sets with different tissue types. I demonstrated the validity of TEA-graph and IG values in lung tissue in the NLST data set. As expected, the TEA-graph showed better risk stratification performance than the TNM step, and each IG value group was enriched in the sorted risk group. And I found well-known pathological features such as low-grade tumor cell proliferation with a lepidic or acinar pattern in the low IG group of lung cancer and polymorphic tumor cell proliferation with a solid pattern and necrosis in the high IG group. All these results

51

showed that the IG values were well indicative of the risk-related significance of the trait and that the TEA-graph adequately extracted the prognostic-related contextual trait.

3.3.3. Validation of interpretability for different event
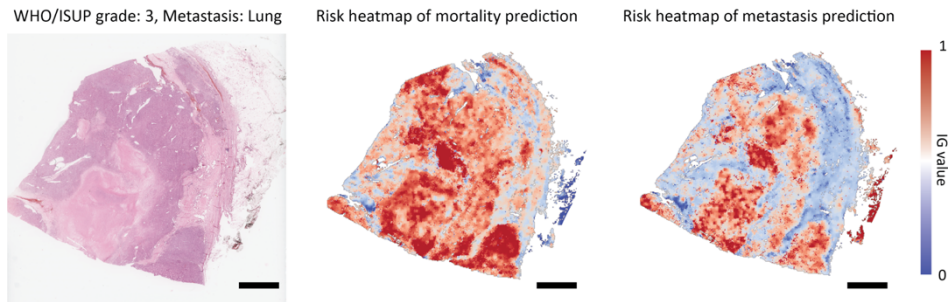
Figure 3.17 Differences between the predicted prognostic features of survival and metastasis.

Since TEA-graph requires only WSI-level annotation in a semi-supervised manner, various prognostic characteristics of survival and metastasis events can be compared. To this end, I trained two separate models on cancer-specific survival and metastasis-free survival data. I used two separate models to obtain risk-related contextual features important for survival and metastatic events. To compare risk-related contextual features between two events, I observed risk-related areas of mortality and metastasis in the same patient using IG values from two separate models. In particular, by extracting the connection graph of the super patch with the top 10% IG value, the prognostic characteristics according to the situation are indicated. I then explored unique and shared contextual features for survival and metastasis events. In both cases, the high IG area showed advanced pathological features, such as striated muscle. In addition, I observed sarcoma and intratumoral lymphocyte infiltration.
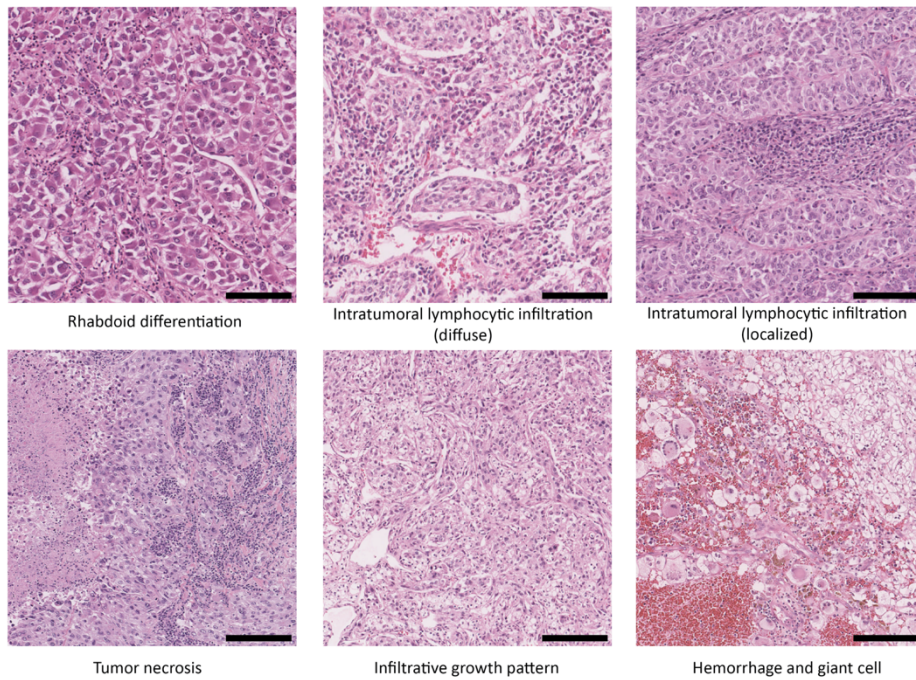
Figure 3.18 Pathological features of the connected graph that had a high IG value and appeared in both survival and metastasis events. Scale bar, 400μm.

Although pathologic features were shared in survival and metastasis, regions that appeared predominantly in the survival events showed tumor necrosis, infiltrative growth or discohesive tumor cells. The difference in the pathological features between the two events means that the TEA-graph learned the different context features according to the target events. Although minimal information about the contextual features of metastasis is available, the high-risk features that appeared in metastasis events can be candidates for metastasis-specific pathological biomarkers.
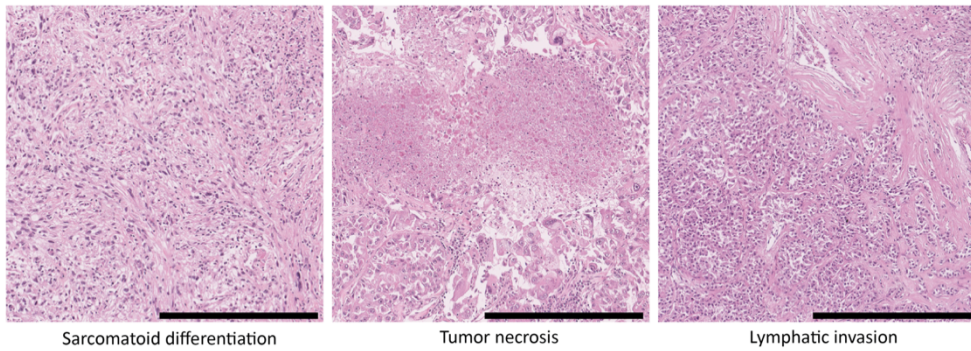
Figure 3.19 Pathological features of the connected graph that had a high IG value and appeared in both survival and metastasis events. Scale bar, 400μm.

3.4.   Contextual pathological biomarker

3.4.1.   Biomarker discovery

1.Clustering the context-feature embedded pathological features

2.Each cluster represents different pathological features

3. Define pathological features with patch-level interaction
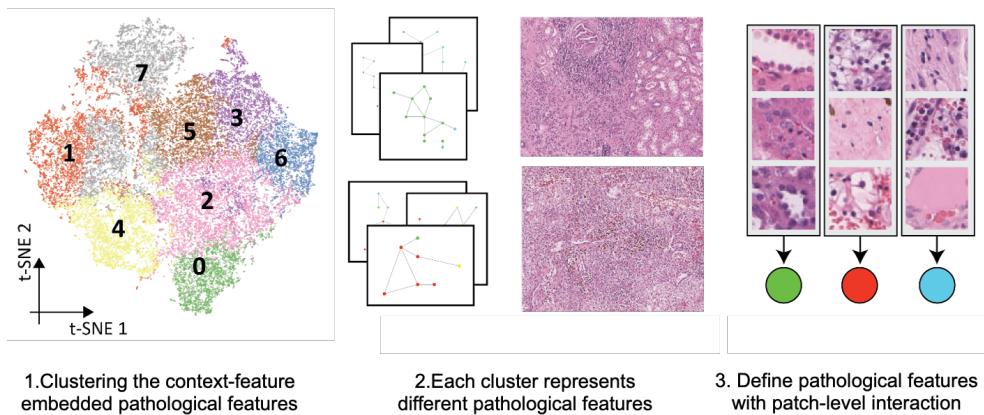
Figure 3.20 Overall workflow to discover the pathological biomarkers that have prognostic power

I investigated whether TEA-graphs extracted useful contextual features to define prognostic biomarkers in a data-driven manner. The IG values of all ccRCC patients were calculated and a subgraph of each IG group, a linked graph (top 10% for top group, bottom 10% for subgroup) was extracted to represent situational features relevant to prognosis. To cluster the subgraphs with specific pathological features, cluster features with 8 different marker groups in the subgraphs of each IG group

using the k-means clustering method and use t-stochastic neighbor embeddings (t-SNE). to visualize the cluster. Then, a Kaplan-Meier survival analysis was performed to measure the difference in the area under the Kaplan-Meier plot between large and small groups that could reflect the risk of that subgraph-level cluster. I observed that the area under the plot was biased towards negative values in the low IG group and biased towards positive values in the high IG group.

Choose the number of graph clusters that represent the statistically strongest patient stratification performance (p-value of the log-rank test) and clustering performance. I first calculated the number of pathologic features corresponding to each graph cluster per patient for the entire data set. High and low IG graph clusters were associated with adverse and favorable pathological features, respectively. Therefore, groups of patients with a number of High and Low IG graph cluster pathological features above the median of the count values for the entire patient group should be able to distinguish those who do not. The Mid IG cluster, on the other hand, is associated with moderate or unrelated pathologic features, so the patient group should be indistinguishable no matter how many or few pathological features each patient has. Based on the above hypothesis, I measured the fraction of well-segmented High and Low IG clusters with p-values lower than 0.01 in the log-rank test of patient stratification.
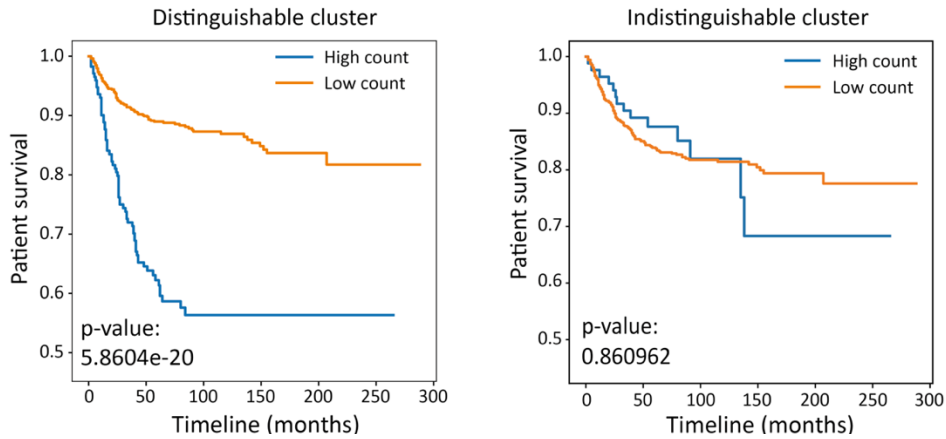
Figure 3.21 Kaplan-Meier plot of the distinguishable and indistinguishable graph cluster

I measured the ratio of the number of distinguishable cluster number, which means the proportion of clusters with a p-value lower than 0.01 in the high group and the low group. As a result, I decided eight as the graph cluster number which showed the highest statistically meaningful cluster ratio in original manuscript.
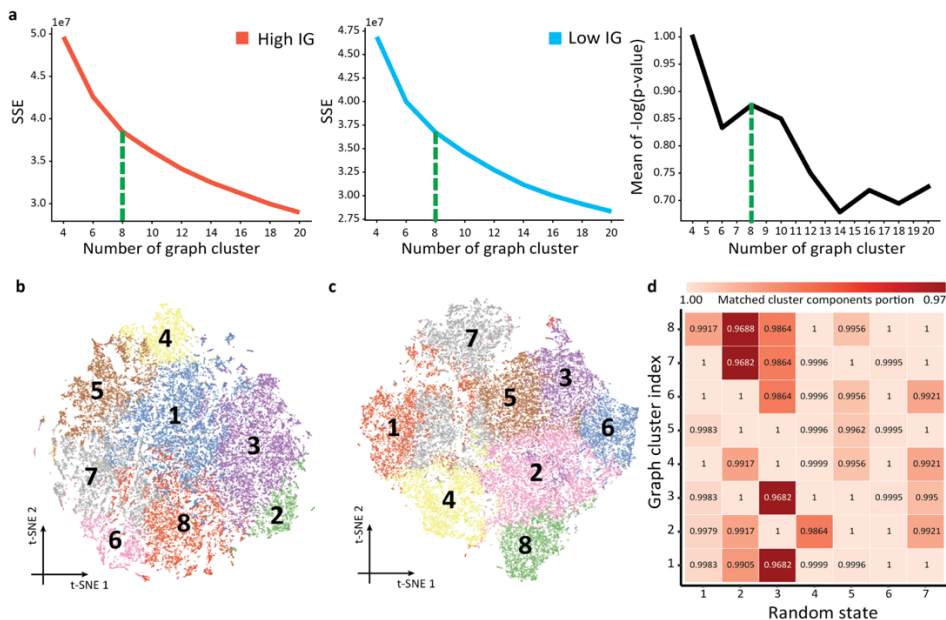
Figure 3.22 a, Sum of square error (SSE) of each cluster in high IG (left) and low IG (right) with different graph cluster number b, t-SNE plot of subgraph clustered by k-means clustering method of high IG. c, t-SNE plot of subgraph clustered by k-means clustering method of low IG. d, Matched portion of subgraph between various randomly initialized clustering results with original clustering result.

56

I first used the K-means clustering method to cluster the graph features and then used the t-SNE method as the visualization and validation method of proposed clustering results. I used the hyperparameters of t-SNE as [number of components = 2, perplexity = 30, iteration number = 1000, learning_rate = 200] and it can clearly visualize K-means clustering results. The clustering method is purely unsupervised method so the different initialization can affect the clustering result. I tested whether different initialization affect a clustering result and confirmed the components that originally included in each graph cluster were preserved even though different initialization. Because the cluster index of randomly initialized clustering can be changed, I first found the appropriate cluster index that the original cluster index is included the most. And then I measured the accuracy of how many components are preserved in the randomly initialized clustering in total of seven different random state. I found that approximately 98.8% of original components were preserved even after multiple randomly initialized clustering, therefore proposed graph clustering method is robust against effect of randomness of unsupervised clustering method. Although the prognostic characteristics of the graph cluster were maintained so that the contextual pathological features could be found on every clustering analysis, the user might obtain slightly different results on every clustering analysis. This is the limitation of unsupervised, clustering-based analysis, and I can optimize the consistency and reliability of the clustering results further for each application, for example, long iterations for high convergence.
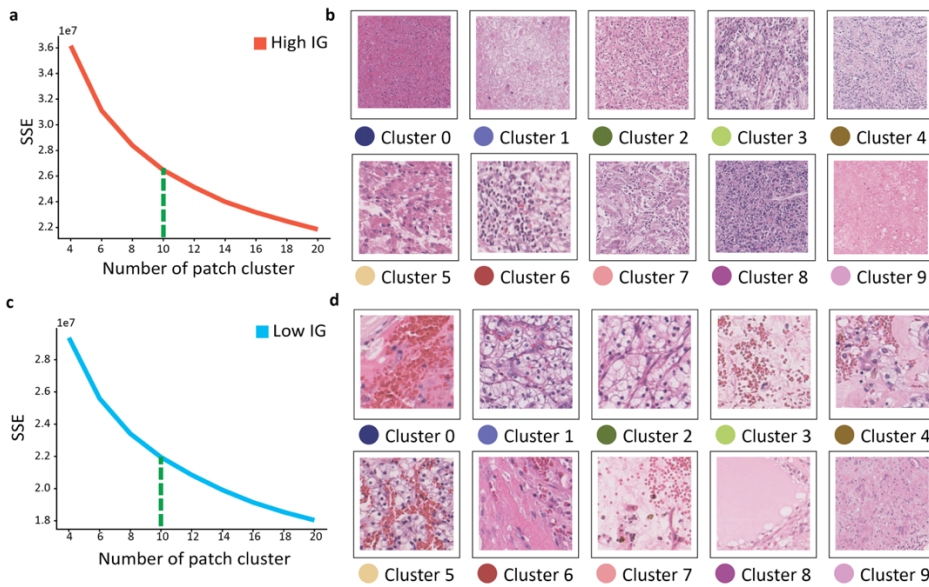
Figure 3.23 a, Sum of square error (SSE) of each cluster in high IG with different patch cluster number. b, Pathological features of each cluster in high IG. c, Sum of square error (SSE) of each cluster in low IG with different patch cluster number. d, Pathological features of each cluster in low IG.

In the patch level clustering, I selected the number of clusters based on the elbow point of the SSE of each patch cluster to better decide the number of clusters. I choose the ten as the number of cluster for the patch level clustering because that is the point where the difference of SSE is linear. However, since the purpose of patch clustering is to interpret the relationship between patches in a contextual pathological feature, there are various options to choose the appropriate patch cluster number if the number satisfies the certain statistical criteria. Therefore, I also discussed the number of patch cluster to represent the pathological features I measured in the subgraph features. Likewise, each user can determine the appropriate patch cluster number according to the level of pathological details that the user wants to measure.

After calculating the IG values for the entire data set, group the top 10%, middle 10%, and bottom 10% IG values as targets for context graph biomarker extraction. A connected graph is defined as a graph with 5 or more nodes among the largest

58

graphs connected to the same group (high, medium, low).

Defines a connected graph feature by averaging all node features in the connected graph. For node features, I connect two features: morphological features from the pre-trained CNN model and graphical features from the trained TEA graph. Choose the graph cluster that represents the most distinguishable clusters. All associated graph features were normalized to the same group (high, medium, low) and k-means clustering was performed to extract each pathological environment biomarker from each group. A t-SNE plot was used to visualize the clustering results. I then counted the number of subgraphs in each graph cluster for each patient and performed a Kaplan-Meier analysis with high and low groups. Higher count groups represent more patients in the subgraphs of the selected graph cluster than the median count values of the selected graph clusters, and the lower count groups represent the opposite. I obtain the Kaplan-Meier plot of each cluster and measure the difference between the areas under the Kaplan-Meier plot of the high and low count groups to rank the importance of each subgroup as prognostic biomarkers (The p-value is Calculated by log-rank test). All survival analysis is performed using Python's Lifeline package. I define clusters with p-values less than 0.01 as distinguishable clusters in the high/low group. Choose the graph cluster number with the highest number of distinguishable clusters in the high/low group.

  For patch-by-patch analysis, I first sampled 100 subgraphs from each graph cluster. Then, patch-level indices were created with k-means clustering using the morphological features of the patches in the sampled subgraphs. In the selected subgraph cluster, I assigned a patch-level cluster index to each node in the subgraph. I then counted each pair of patch-level cluster indices connected by the edges of the subgraph and normalized to the total number of edges in the subgraph. From the

59

subgraph clusters, patch clusters exhibiting high connectivity were selected and the pathological features of pairs of patch-level clusters were interpreted based on the pathological features of each patch-level cluster. I labeled nodes with superpatch labels to visualize each connected graph using networkx and to characterize the label connection patterns of each connected graph.

To determine whether the selected graph features had prognostic features, I extracted all linked graph features that were in the same cluster and exhibited high correlation (> 0.9) with the selected graph features. Then, using the count matrix of the highly correlated subgraph, a Kaplan-Meier plot was obtained and the p-value was calculated.

Graphical analysis was performed to extract the contextual characteristics of each cluster. To this end, I define patch-level pathologic features grouped into 10 different pathological features from the high IG group and the low IG group, respectively. Each patch-level cluster exhibits different morphological features depending on the IG group, such as necrosis and intratumoral lymphocyte infiltration for the high IG group, vitrification and cystic changes in the low IG group. Second, I identified patch connectivity patterns in each subgraph cluster. Each subgraph cluster has its own patch connection pattern, and the cluster pattern can be determined by a heatmap. For example, cluster 6 of the high-risk graph in the high IG group shows patterns of abscess formation and intratumoral lymphocyte infiltration (patch cluster 3).
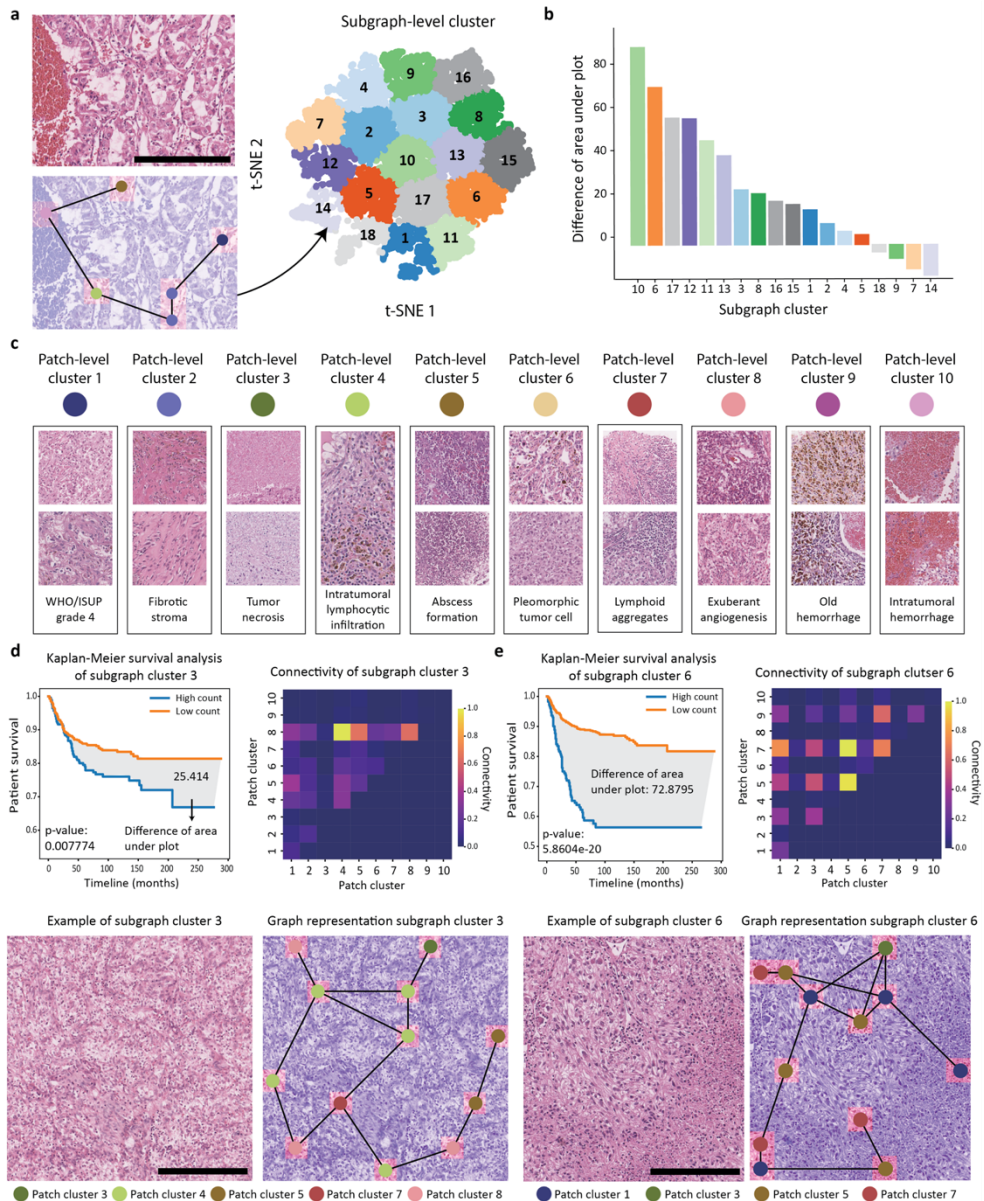
Figure 3.24 Risk-related environmental markers predicted by TEA-graph. **a,**
Example of the subgraph features that is the input data for the t-SNE analysis (left),
subgraph-level clustering using t-SNE and kNN (right). Scale bar, 256μm. **b,** Plot
of the difference of area under plot betIen low count and high count Kaplan–Meier
plot which reflects the risk of the clusters. **c,** Patch level cluster characteristic of the
high IG group. **d, e,** Kaplan–Meier plot and edge distribution (connectivity) of
subgraph-level cluster 3, 6. Connectivity shows which patch clusters are interacted
with each other more frequently (top). Example of the subgraph in subgraph-level
cluster 3, 6. Each node color indicate the corresponding patch cluster (bottom).
Scale bar, 200μm.
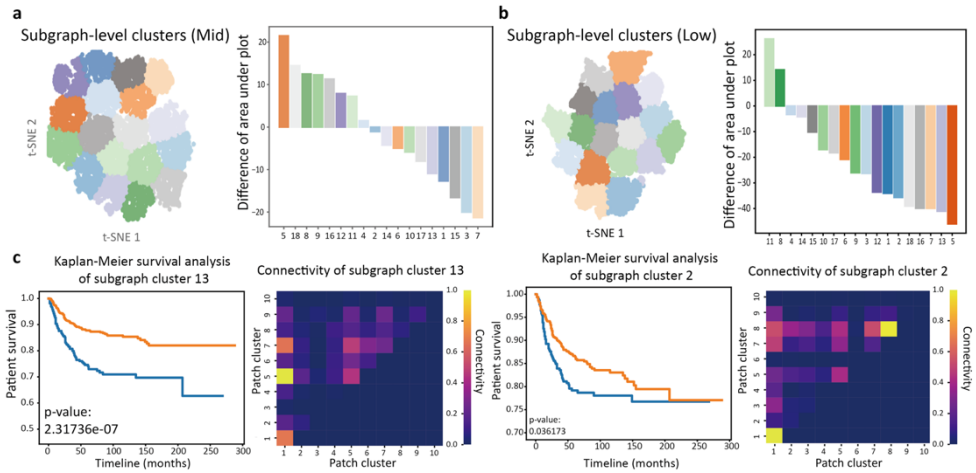
Figure 3.25 More examples of risk-related environmental markers predicted by TEA-graph. **a, b,** t-SNE and the difference of area under plot of mid IG group and low IG group. **c,** Kaplan–Meier plot and connectivity of subgraph cluster 13, 2 from top IG group (left, right). Connectivity shows inter-connected frequency of each patch cluster.

This contextual feature was well captured because, unlike other tumor types, lymphocyte infiltration presents an adverse prognostic marker that is a unique feature of ccRCC and TEA-graph [53]–[55]. To further indicate that each graph cluster exhibits a specific contextual function, a subgraph well representing the pathological features of graph cluster 6 mentioned above was chosen. I then identified other subgraphs that were well clustered in the same 6 graph clusters with pathologically similar features. In graph cluster 6, I found several subgraphs with similar pathological features and identified an unfavorable prognosis in those subgraphs..

Additionally, graph cluster 3 of the low IG group shows stroma vitrification (patch cluster 4) associated with low-grade tumor cells (patch cluster 2) and chicken wire vasculature (patch cluster 7). As with the group with high IG, a representative subgraph was selected from graph cluster 3, and a good prognosis was confirmed

with a subgraph with similar pathological characteristics in graph cluster 3.

3.4.2. Proposed new pathological marker

Interestingly, I found angiogenesis-related pathological features in cluster 0 of the high IG graph, called active granulation and angiogenic focus (AGAF). I found that patch clusters 6 with active granular and angiogenic functions were correlated with patch clusters 3 and 5, respectively, which were associated with intratumoral lymphocyte infiltration and eosinophilic/transformation functions.
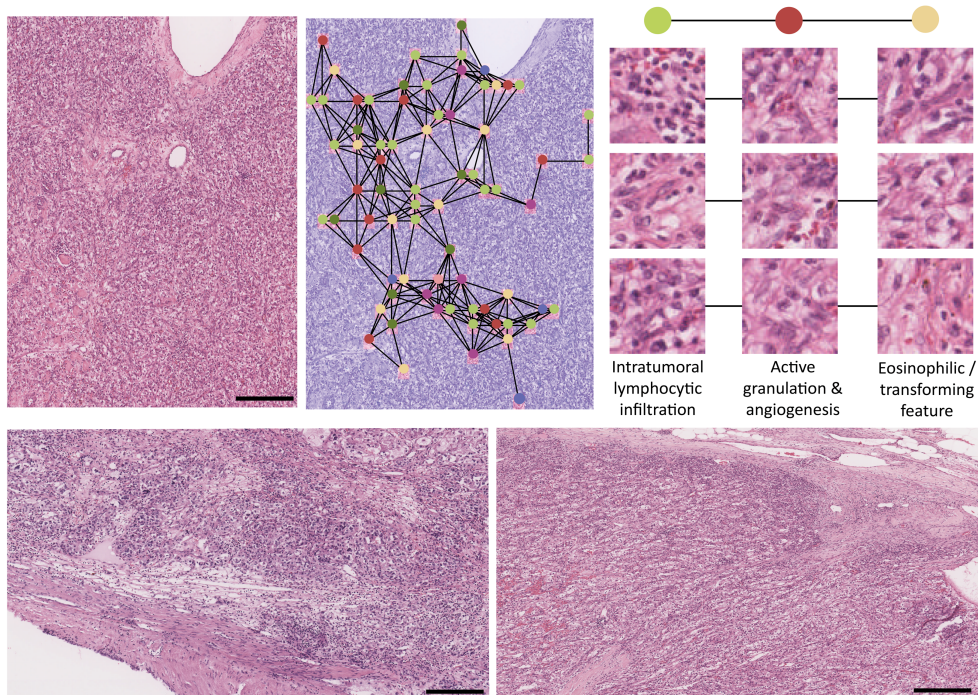
Figure 3.26 Angiogenesis-related environmental markers predicted by the TEA-graph

Cancer progression alters the tumor microenvironment towards more vascularization, and fibroblast proliferation and immune cell infiltration can reveal granulation tissue formation and form cancer-prone niches (Patch cluster 6) [56]–[58]. In this case, cancer cells can interact closely with fibroblasts, immune cells and blood vessels

(patch cluster 5), which can lead to intratumoral lymphocyte infiltration (patch cluster 3). Thus, constellation of AGAF, eosinophilic/transformation characteristics, and intratumoral lymphocyte infiltration have been associated with disease progression and poor prognosis. I selected representative subgraphs of AGAF and, as expected, found several subgraphs with AGAF features in graph cluster 0. Overall, I confirmed that TEA-graphs identified histopathological prognostic features in a semi-supervised and data-driven manner using substantial WSI data sets.

In chapter 4, I will discuss the limitation of proposed method and room for improvement that combined with recently developed method.

Chapter 4.

Conclusion and discussion

TEA-graph show promising results to predict the patient's prognosis incorporating context features. Because TEA-graph is end-to-end framework consist of multiple components, it's important to find out which component is critical for improved performance of TEA-graph. Also, though TEA-graph show promising results with incorporating context features, there is still a room for improvement. In this chapter, There are multiple components within TEA-graph there is still a room for improvement. In this chapter, hyperparameter screening and ablation study of TEA-graph is proposed and also discussed recent development of graph neural network that could further improve the TEA-graph.

I showed that TEA-graphs represent contextual features of WSI and can learn prognostic features, including complex interactions with the tumor environment. TEA-Graph has demonstrated improved risk prediction and stratification performance that can help both patients and clinicians in clinical settings. In particular, I used the interpretive method to address the meaningful histopathological features that the TEA-graph has learned to discriminate among patients' risk values. By using the contextual features extracted through TEA-graph, I can cover several contextual features related to the prognosis of ccRCC. I have demonstrated that findings are consistent with previously reported contextual features of ccRCC. The TEA-graph consists of various hyperparameters optimized to adequately predict prognostic characteristics. I found that utilizing edge features to represent the relative positions (distance and angle) between superpatches had a decisive effect on the performance of the TEA-graph.

a

Positional	Angle	Embedding	Accuracy
No	No	No	0.81857
Yes	No	No	0.82537
No	Yes	No	0.82871
Yes	No	Yes	0.83631
No	Yes	Yes	0.84260
Yes	Yes	Yes	**0.84832**

b

Correlation	Planar graph	Required RAM	Accuracy
0.5	No	0.5x	0.83082
0.65	No	0.7x	0.8309
0.75	Yes	0.9x	0.82197
0.75	No	1x	**0.84832**

Table 4 Ablation study of different edge features used for TEA-graph

Simple multiplications on node features also benefit performance, but you can achieve better performance by including distance and angle features as learnable parameters. These results are consistent with the results of other studies that utilize location information to better discriminate neighboring features during aggregation via GNN [59], [60]. In addition, in order to secure optimal performance, the method of constructing a graph is important. Unlike typical Euclidean space-by-patch graphs where the distances between each patch are the same, the super patch graph can have different edge states because the distances between each patch are different. I

achieved better performance when I allowed intersections between edges, which means that the constructed graph is not planar. A ~2% performance improvement over planar graphs was obtained when edges were created between all patches at less than 3 patch distances.

a

Tuning parameters	Search space	Selected value	Tuning parameters	Search space	Selected value
GNN model	[GAT, GIN, EdgeConv]	GAT	Normalization	[BatchNorm, LayerNorm]	LayerNorm
Pretrained CNN	[ImageNet, T/N classifier]	T/N classifier	Activation	[ReLU, LeakyReLU, PReLU]	PReLU
GNN layer	[1, 2, 3]	3	Dropout	[0, 0.25, 0.5]	0.25
MLP layer	[2, 3, 4]	3	Dropedge	[0, 0.25, 0.5]	0.25
Residual	[Yes, No]	Yes	Dropfeature	[0, 0.25, 0.5]	0.25

b

Model	Pretrain	GNN layers	MLP Layers	Residual	Norm	Activation	Dropout	Dropedge	Dropfeature	Accuracy
GIN	T/N classifier	3	3	Yes	LayerNorm	PReLU	0.25	0.25	0.25	0.8244
EdgeConv	-	-	-	-	-	-	-	-	-	0.7848
GAT	ImageNet	-	-	-	-	-	-	-	-	0.8166
-	T/N classifier	2	-	-	-	-	-	-	-	0.8427
-	-	1	-	-	-	-	-	-	-	0.8358
-	-	3	4	-	-	-	-	-	-	0.8420
-	-	-	2	-	-	-	-	-	-	0.8464
-	-	-	3	No	-	-	-	-	-	0.8317
-	-	-	-	Yes	BatchNorm	-	-	-	-	0.8129
-	-	-	-	-	LayerNorm	-	-	-	-	0.8483
-	-	-	-	-	-	ReLU	-	-	-	0.8451
-	-	-	-	-	-	LeakyReLU	-	-	-	0.8466
-	-	-	-	-	-	PReLU	-	-	-	0.8483
-	-	-	-	-	-	-	0.0	0.0	0.0	0.8240
-	-	-	-	-	-	-	0.25	-	-	0.8412
-	-	-	-	-	-	-	0.5	-	-	0.8238
-	-	-	-	-	-	-	0.25	0.0	-	0.8323
-	-	-	-	-	-	-	-	0.25	-	0.8429
-	-	-	-	-	-	-	-	0.5	-	0.8205
-	-	-	-	-	-	-	-	0.25	0.5	0.8287
GAT	T/N classifier	3	3	Yes	LayerNorm	PReLU	0.25	0.25	0.25	**0.8483**

c

GNN model	GIN	EdgeConv	GAT	Std	Normalization	BatchNorm		LayerNorm	Std
	0.8244	0.7848	0.8483	0.032		0.8129		0.8483	0.025

Pretrained CNN	ImageNet	T/N classifier	Std	Activation	ReLU	LeakyReLU	PReLU	Std
	0.8166	0.8483	0.022		0.8451	0.8466	0.8483	0.002

GNN layer	1	2	3	Std	Dropout	0	0.25	0.5	Std
	0.8358	0.8457	0.8483	0.006		0.8240	0.8412	0.8238	0.014

MLP layer	2	3	4	Std	Dropedge	0	0.25	0.5	Std
	0.8464	0.8483	0.8420	0.003		0.8323	0.8429	0.8205	0.008

Residual	Yes	No	Std	Dropfeature	0	0.25	0.5	Std
	0.8483	0.8317	0.012		0.8240	0.8483	0.8287	0.014

Table 5 Hyperparameter screening of TEA-graph

TEA-graph also introduces a superpatch approach to compress WSI without loss of context function, alleviating the hardware requirements of multiple GPUs due to the gigapixel size of WSI. Collectively, proposed approach proposes to use modified GATs to learn contextual features of WSIs in a semi-supervised manner while processing all WSIs more efficiently. There is a trade-off between accuracy and compression ratios that change based on the correlation threshold used to construct

the superpatch. The best accuracy was achieved with a correlation threshold of 0.75 and a compression ratio of about 50%, but similar results were achieved with a correlation threshold of 0.5 and a compression ratio of about 10%. Thus, researchers and clinicians can easily analyze interpretable contextual features with their own GPUs and WSIs using TEA graphs by selecting the optimal parameters for their hardware setup.

I also thoroughly investigated the sensitivity of the TEA-graph to several hyperparameters. Similar results can be achieved by extracting important features from small patches using ImageNet pretrained CNNs, but transfer-training the CNN model to tumor and normal classification can increase performance by ~2%. Also, choosing an appropriate regularization method had a significant impact on the performance of the TEA graph. GNNs are well-known for the problem of over-smoothing, as they easily overfit to data sets due to the nature of the way messages are passed. Due to the large size of the WSI, I could not test deeper GNNs due to out of memory issues, but I achieved better performance when I increased the GNN layers using dropout, dropedge, and node/edge feature dropout (Dropfeature) to alleviate overfitting [61], [62]. Overall, these parameters are critical to achieving the optimal performance with TEA-graph.

Although TEA-graphs can be used to find meaningful contextual biomarkers, there is still room for improvement to more accurately reflect pathological features in contextual features. I selected the most robust graph clusters of the low and high IG groups and found that the TEA-graph confuses homogeneous pathological features with detailed cellular-level features occupying large areas. For example, low-level tumors and lymphocytes sometimes have very similar pathological appearances that are indistinguishable from patch-level features. When vitrification features are too

large to be covered by a three-layer TEA-graph, the TEA-graph sometimes does not match the features with a similar pattern, such as ischemic necrosis features. These limitations can be improved by combining cellular and patch level features [63], [64].
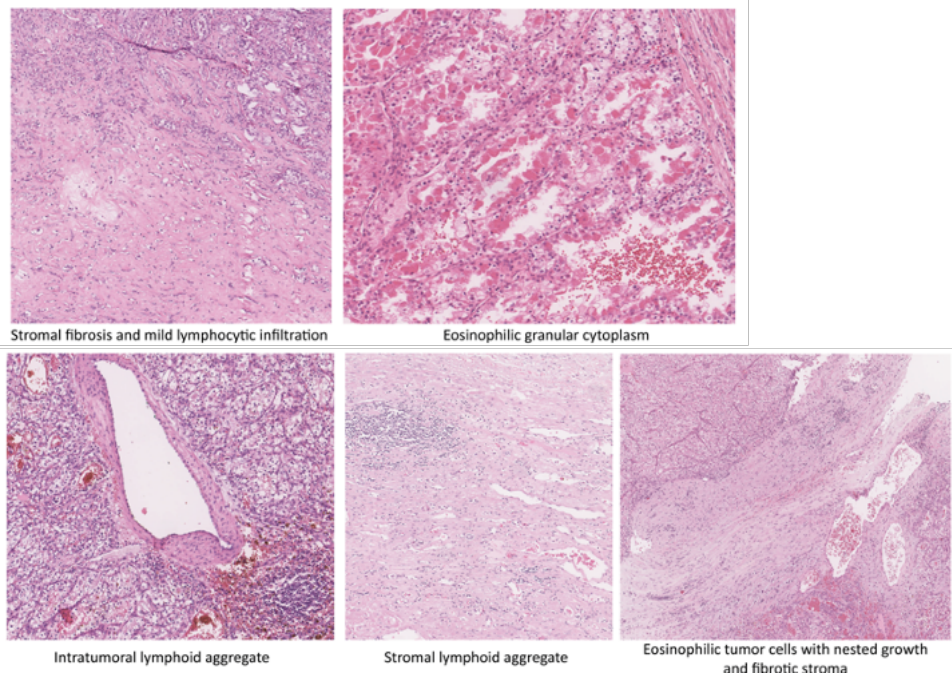
Figure 4.1 Pathological features of each IG group misclassified by TEA-graph.

Additionally, recent GNN models were further developed with attention models that consider the entire WSI correlation without limitation in the stacked layer [65], [66]. Therefore, the risk prediction and contextual biomarker detection performance of the TEA-graph can be further improved through the application of the above-mentioned methods. In addition, a methodology for finding graphic biomarkers with TEA-graph was proposed. I used unsupervised clustering methods such as k-means clustering and t-SNE with subgraph features of GNN. Due to the nature of the Unsupervised Clustering method, an optimization process is required according to the purpose or characteristics of the organization.

Deep learning has facilitated innovation in pathology representing heterogeneous

image data with different interactions of different tumor cells and microenvironments. Numerous studies have demonstrated the utility of deep learning for pathological imaging, but the clinical deployment of these techniques remains an open question [67]. Currently, several clinical studies are being performed that will show the utility of deep learning in aiding pathologists in actual clinical settings [68]–[70]. I believe that proposed study can also contribute to the realization of the clinical benefits of deep learning through contextual feature-based biomarker retrieval in the near future.

Bibliography

[1] K. Vinay, A. K Abul, F. Nelson, and A. C Jon, *Robbins and Cotran pathologic basis of disease, professional edition e-book*. Elsevier health sciences.

[2] E. I Jonathan, A. J. C William, A. B Mahul, E. L Lars, and I. G. Committee, "The 2005 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma," vol. 29, no. 9, pp. 1228–1242.

[3] D. Brett *et al.*, "The International Society of Urological Pathology (ISUP) grading system for renal cell carcinoma and other prognostic parameters," vol. 37, no. 10, pp. 1490–1504.

[4] W. Y Anne and H. David, "WHO/ISUP classification, grading and pathological staging of renal cell carcinoma: standards and controversies," vol. 36, no. 12, pp. 1913–1926.

[5] C. Qi *et al.*, "Ontological analyses reveal clinically-significant clear cell renal cell carcinoma subtypes with convergent evolutionary trajectories into an aggressive type," vol. 51, p. 102526.

[6] W. Yinhai, W. E Kate, K. J Paul, J. A Jacqueline, and H. W Peter, "SurfaceSlide: a multitouch digital pathology platform," vol. 7, no. 1, p. e30783.

[7] Y. Yinyin *et al.*, "Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling," vol. 4, no. 157, pp. 157ra143-157ra143.

[8] B. H Andrew *et al.*, "Systematic analysis of breast cancer morphology uncovers stromal features associated with survival," vol. 3, no. 108, pp. 108ra113-108ra113.

[9] D. Faraggi and R. Simon, "A neural network model for survival data," vol. 14, no. 1, pp. 73–82.

[10] H. Kvamme, Ø. Borgan, and I. Scheel, "Time-to-event prediction with neural networks and Cox regression".

[11] E. Giunchiglia, A. Nemchenko, and M. van der Schaar, "RNN-SURV: A deep recurrent model for survival analysis," pp. 23–32.

[12] P. Mobadersany *et al.*, "Predicting cancer outcomes from histology and genomics using convolutional networks," vol. 115, no. 13, pp. E2970–E2979.

[13] P. Courtiol *et al.*, "Deep learning-based classification of mesothelioma improves prediction of patient outcome," vol. 25, no. 10, pp. 1519–1525.

[14] G. Quellec, G. Cazuguel, B. Cochener, and M. Lamard, "Multiple-instance learning for medical image and video analysis," vol. 10, pp. 213–234.

[15] X. Zhu, J. Yao, F. Zhu, and J. Huang, "Wsisa: Making survival prediction from whole slide histopathological images," pp. 7234–7242.

[16] M. Y. Lu, D. F. Williamson, T. Y. Chen, R. J. Chen, M. Barbieri, and F. Mahmood, "Data-efficient and weakly supervised computational pathology on whole-slide images," vol. 5, no. 6, pp. 555–570.

[17] G. Campanella *et al.*, "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," vol. 25, no. 8, pp. 1301–1309.

[18] O. Ciga, T. Xu, S. Nofech-Mozes, S. Noy, F.-I. Lu, and A. L. Martel, "Overcoming the limitations of patch-based learning to detect cancer in whole slide images," vol. 11, no. 1, pp. 1–10.

[19] D. Tellez *et al.*, "Extending unsupervised neural image compression with supervised multitask learning," pp. 770–783.

[20] L. J. Medeiros, A. B. Gelb, and L. M. Weiss, "Renal cell carcinoma. Prognostic significance of morphologic parameters in 121 cases," *Cancer*, vol. 61, no. 8, pp. 1639–1651, 1988, doi: 10.1002/1097-0142(19880415)61:8<1639::aid-cncr2820610823>3.0.co;2-f.

[21] B. Delahunt *et al.*, "The International Society of Urological Pathology (ISUP) Grading System for Renal Cell Carcinoma and Other Prognostic Parameters," *Am J Surg Pathology*, vol. 37, no. 10, pp. 1490–1504, 2013, doi: 10.1097/pas.0b013e318299f0fb.

[22] B. Sobottka, A. Lorch, K. Silina, M. van den Broek, and H. Moch, "Renal cell carcinoma pathology in 2021: 'new need for renal cancer immune profiling,'" *Curr Opin Urol*, vol. 31, no. 3, pp. 228–235, 2021, doi: 10.1097/mou.0000000000000864.

[23] W. Lu, S. Graham, M. Bilal, N. Rajpoot, and F. Minhas, "Capturing Cellular Topology in Multi-Gigapixel Pathology Images," *CVPRW*, vol. 00, pp. 1049–1058, 2020, doi: 10.1109/cvprw50498.2020.00138.

[24] P. Pati *et al.*, "HACT-Net: A Hierarchical Cell-to-Tissue Graph Neural Network for Histopathological Image Classification," *Arxiv*, 2020.

[25] R. J. Chen *et al.*, "Pathomic Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features for Cancer Diagnosis and Prognosis," *IEEE Transactions on Medical Imaging*, 2019.

[26] Y. Zhao *et al.*, "Predicting Lymph Node Metastasis Using Histopathological Images Based on Multiple Instance Learning with Deep Graph Convolution," *CVPR*, vol. 00, pp. 4836–4845, 2020, doi: 10.1109/cvpr42600.2020.00489.

[27] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," pp. 701–710.

[28] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the web.," Stanford InfoLab.

[29] P. W. Battaglia *et al.*, "Relational inductive biases, deep learning, and graph networks".

[30] R. Li, J. Yao, X. Zhu, Y. Li, and J. Huang, "Graph CNN for survival analysis on whole slide pathological images," pp. 174–182.

[31] W. Lu, M. Toss, M. Dawood, E. Rakha, N. Rajpoot, and F. Minhas, "SlideGraph+: Whole Slide Image Level Graphs to Predict HER2 Status in Breast Cancer," p. 102486.

[32] R. J. Chen *et al.*, "Whole Slide Images are 2D Point Clouds: Context-Aware Survival Prediction using Patch-based Graph Convolutional Networks," pp. 339–349.

[33] M. T. Ribeiro, S. Singh, and C. Guestrin, "' Why should i trust you?' Explaining the predictions of any classifier," pp. 1135–1144.

[34] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," vol. 30.

[35] Z. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec, "Gnnexplainer: Generating explanations for graph neural networks," vol. 32.

[36] B. Sanchez-Lengeling *et al.*, "Evaluating attribution for graph neural networks," vol. 33, pp. 5898–5910.

[37] S. Fortunato, "Community detection in graphs," *Phys Reports*, vol. 486, no. 3–5, pp. 75–174, 2010, doi: 10.1016/j.physrep.2009.11.002.

[38] N. Stanley, R. Kwitt, M. Niethammer, and P. J. Mucha, "Compressing Networks with Super Nodes," *Sci Rep-uk*, vol. 8, no. 1, p. 10892, 2018, doi: 10.1038/s41598-018-29174-3.

[39] B. Gary, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, vol. 3, 2000.

[40] G. Adam, G. Benjamin, H. Jan, J. Drazen, and S. Mahadev, "OpenSlide: A vendor-neutral software foundation for digital pathology," *J Pathol Inform*, vol. 4, no. 1, pp. 27–27, 2013, doi: 10.4103/2153-3539.119005.

[41] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph Attention Networks," *ICLR*, 2017.

[42] V. P. Dwivedi, C. K. Joshi, T. Laurent, Y. Bengio, and X. Bresson, "Benchmarking Graph Neural Networks," *ArXiv*, vol. abs/2003.00982, 2020.

[43] J. You, Z. Ying, and J. Leskovec, "Design Space for Graph Neural Networks," *NeurIPS*, vol. 33, pp. 17009–17021, 2020, [Online]. Available: https://proceedings.neurips.cc/paper/2020/file/c5c3d4fe6b2cc463c7d7ecba17cc9de7-Paper.pdf

[44] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *PMLR*, vol. 37, pp. 448–456, Jun. 2015, [Online]. Available: http://proceedings.mlr.press/v37/ioffe15.html

[45] M. Fey and J. E. Lenssen, "Fast Graph Representation Learning with PyTorch Geometric," *ArXiv*, vol. abs/1903.02428, 2019.

[46] N. Kokhlikyan *et al.*, "Captum: A unified and generic model interpretability library for PyTorch," *Arxiv*, 2020.

[47] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic Attribution for Deep Networks," *PMLR*, 2017.

[48] W. B Alexander *et al.*, "Evaluating Attribution for Graph Neural Networks," *Advances in neural information processing systems*, 2020.

[49] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov, and A. Smola, "Deep Sets," *Arxiv*, 2017.

[50] R. J. Chen *et al.*, "Whole Slide Images are 2D Point Clouds: Context-Aware Survival Prediction using Patch-based Graph Convolutional Networks," *Arxiv*, 2021.

[51] M. Ilse, J. M. Tomczak, and M. Welling, "Attention-based Deep Multiple Instance Learning," *Arxiv*, 2018.

[52] C. Davidson-Pilon, "lifelines: survival analysis in Python," *J Open Source Softw*, vol. 4, no. 40, p. 1317, 2019, doi: 10.21105/joss.01317.

[53] S. Zhang *et al.*, "Immune infiltration in renal cell carcinoma," *Cancer Sci*, vol. 110, no. 5, pp. 1564–1572, 2019, doi: 10.1111/cas.13996.

[54] D. A. Braun *et al.*, "Interplay of somatic alterations and immune infiltration modulates response to PD-1 blockade in advanced clear cell renal cell carcinoma," *Nat Med*, vol. 26, no. 6, pp. 909–918, 2020, doi: 10.1038/s41591-020-0839-y.

[55] Y. Qi *et al.*, "Tumor-infiltrating CD39+CD8+ T cells determine poor prognosis and immune evasion in clear cell renal cell carcinoma patients," *Cancer Immunol Immunother*, vol. 69, no. 8, pp. 1565–1576, 2020, doi: 10.1007/s00262-020-02563-2.

[56] A. A. Hakimi *et al.*, "An Integrated Metabolic Atlas of Clear Cell Renal Cell Carcinoma," *Cancer Cell*, vol. 29, no. 1, pp. 104–116, 2016, doi: 10.1016/j.ccell.2015.12.004.

[57] M. Tjota, H. Chen, M. Parilla, P. Wanjari, J. Segal, and T. Antic, "Eosinophilic Renal Cell Tumors With a TSC and MTOR Gene Mutations Are Morphologically and Immunohistochemically Heterogenous," *Am J Surg Pathology*, vol. 44, no. 7, pp. 943–954, 2020, doi: 10.1097/pas.0000000000001457.

[58] W. K. Rathmell, J. C. Rathmell, and W. M. Linehan, "Metabolic Pathways in Kidney Cancer: Current Therapies and Future Directions," *J Clin Oncol*, vol. 36, no. 36, pp. 3540–3546, 2018, doi: 10.1200/jco.2018.79.2309.

[59] J. You, R. Ying, and J. Leskovec, "Position-aware Graph Neural Networks," 2019.

[60] H. Xu, C. Jiang, X. Liang, and Z. Li, "Spatial-aware Graph Relation Network for Large-scale Object Detection," *2019 Ieee Cvf Conf Comput Vis Pattern Recognit Cvpr*, vol. 00, pp. 9290–9299, 2019, doi: 10.1109/cvpr.2019.00952.

[61] Y. Rong, W. Huang, T. Xu, and J. Huang, "DropEdge: Towards Deep Graph Convolutional Networks on Node Classification," *Arxiv*, 2019.

[62] Y. You, T. Chen, Y. Sui, T. Chen, Z. Wang, and Y. Shen, "Graph Contrastive Learning with Augmentations," *Arxiv*, 2020.

[63] W. Lu, M. Toss, E. Rakha, N. Rajpoot, and F. Minhas, "SlideGraph+: Whole Slide Image Level Graphs to Predict HER2Status in Breast Cancer," *Arxiv*, 2021.

[64] G. Jaume *et al.*, "Quantifying Explainers of Graph Neural Networks in Computational Pathology," *2021 Ieee Cvf Conf Comput Vis Pattern Recognit Cvpr*, vol. 00, pp. 8102–8112, 2021, doi: 10.1109/cvpr46437.2021.00801.

[65] C. Ying *et al.*, "Do Transformers Really Perform Bad for Graph Representation?," *Arxiv*, 2021.

[66] V. P. Dwivedi and X. Bresson, "A Generalization of Transformer Networks to Graphs," *Arxiv*, 2020.

[67] A. Esteva *et al.*, "Deep learning-enabled medical computer vision," *Npj Digital Medicine*, vol. 4, no. 1, p. 5, 2021, doi: 10.1038/s41746-020-00376-2.

[68] O.-J. Skrede *et al.*, "Deep learning for prediction of colorectal cancer outcome: a discovery and validation study," *Lancet*, vol. 395, no. 10221, pp. 350–360, 2020, doi: 10.1016/s0140-6736(19)32998-8.

[69] R. Yamashita *et al.*, "Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study," *Lancet Oncol*, vol. 22, no. 1, pp. 132–141, 2021, doi: 10.1016/s1470-2045(20)30535-0.

[70] W. Bulten *et al.*, "Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study," *Lancet Oncol*, vol. 21, no. 2, pp. 233–241, 2020, doi: 10.1016/s1470-2045(19)30739-9.

국문 초록

본 연구는 WSI의 (WSI, Whole slide image) 메모리 효율적인 그래프 표현과 맥락적 (Contextual) 특징을 고려하는 해석 가능한 그래프 신경망 (GNN, Graph neural network)을 기반으로 반 감독 방식 (Semi-supervised learning)의 그래프 학습을 사용하는 종양 환경 관련 맥락 분석이라는 새로운 접근 방식인 TEA-graph를 제안합니다. 디지털 스캔한 종양 단면인 기가픽셀 해상도의 WSI는 양적 및 데이터 기반 방식으로 병리학 분석의 접근 방식을 변경했습니다. 딥 러닝과 WSI의 결합은 임상적으로 중요한 작업에서 더 많은 시너지 효과를 가져오는 동시에 병리학적 분석의 패러다임을 바꾸었습니다. WSI에 대한 딥 러닝은 CAD (Computer-Aided diagnosis) 및 자동화된 분석을 위한 보조 도구로서 유례없는 결과를 보여주었지만 여전히 "환경적 병리학적 특징"인 종양 미세 환경 (TME, tumor microenvironment)과 같은 임상적으로 중요한 병리학적 특징을 설명할 수 없었습니다.

본 연구에서 제안하는 TEA-graph는 메모리 효율적이고 반 감독 방식으로 WSI의 "환경적 병리학적 특징"을 학습하는 방식입니다. 기존에도 맥락 학습을 위해 그래프 신경망을 사용하고 다양한 방식으로 전체 슬라이드 이미지 (WSI)를 압축한다는 유사한 개념이 도입된 바 있지만, 다양한 종양 미세 환경을 구별하는 중요한 특징인 기하 구조적 특징을 그래프 신경망에 도입한 연구는 없었습니다. 본 연구는 제안한 슈퍼패치 (Superpatch)라는 네트워크 압축 기술을 통해 WSI에 대한 맥락적 정보를 잃지 않고 한번에 WSI를 활용하였습니다. 따라서 TEA-graph는 전문가의 모니터링이 없어도 대규모 WSI

데이터셋으로 확장이 가능하고 연구자와 임상의가 해석할 수 있는 정보까지 제공할 수 있습니다. 제안한 기술의 검증을 위해 4가지 다른 유형의 암으로 구성된 총 4,967명의 환자로부터 얻은 WSI 데이터를 사용하여 TEA-graph 의 예후 예측 성능을 확인했습니다. 특히, TEA-graph는 1366명의 환자로 부터 얻은 데이터 세트에서 투명 세포 신세포 암종(Clear cell renal cell carcinoma, ccRCC)의 이질적 맥락 특징을 포착하고 이러한 특징을 반영하 여 ccRCC의 예후를 예측했습니다. 또한 TEA-graph는 해석 가능한 맥락적 특징을 제공하고 이를 서로 다른 예후 효과를 갖는 여러 그룹으로 분류할 수 있게 해주었습니다. 이와 같은 결과는 TEA-graph가 조직병리학적 특징의 복잡한 상호작용을 식별하여 데이터 기반 방식으로 예후 바이오마커를 제공 했음을 나타냅니다. 또한, ccRCC의 새로운 예후 바이오마커적 특징을 제안 하였으며 성능 비교를 포함한 하이퍼파라미터 스크리닝 (Hyperparameter screening)을 통해 최적의 성능을 위해 TEA-graph의 파이프라인 상에서 중요한 부분을 확인했습니다. 본 연구는 병리학적 조직에 맥락적 특징 학습 을 활용하였지만, 이는 MRI, CT 등 다양한 데이터 종류에 활용될 수 있습니 다. 맥락적 특징 학습은 임상의에게 새로운 통찰력을 제공하고 병리학 분석 모델의 전반적인 성능을 향상시킬 것이기 때문에 향후 딥러닝을 의료 응용 에 활용하는 다양한 분야의 연구자에게 도움이 되고 활용 될 것으로 기대됩 니다.

주요어 : 그래프 신경망, 병리학, 생존 분석, 암 미세환경
학번 : 2016-23108