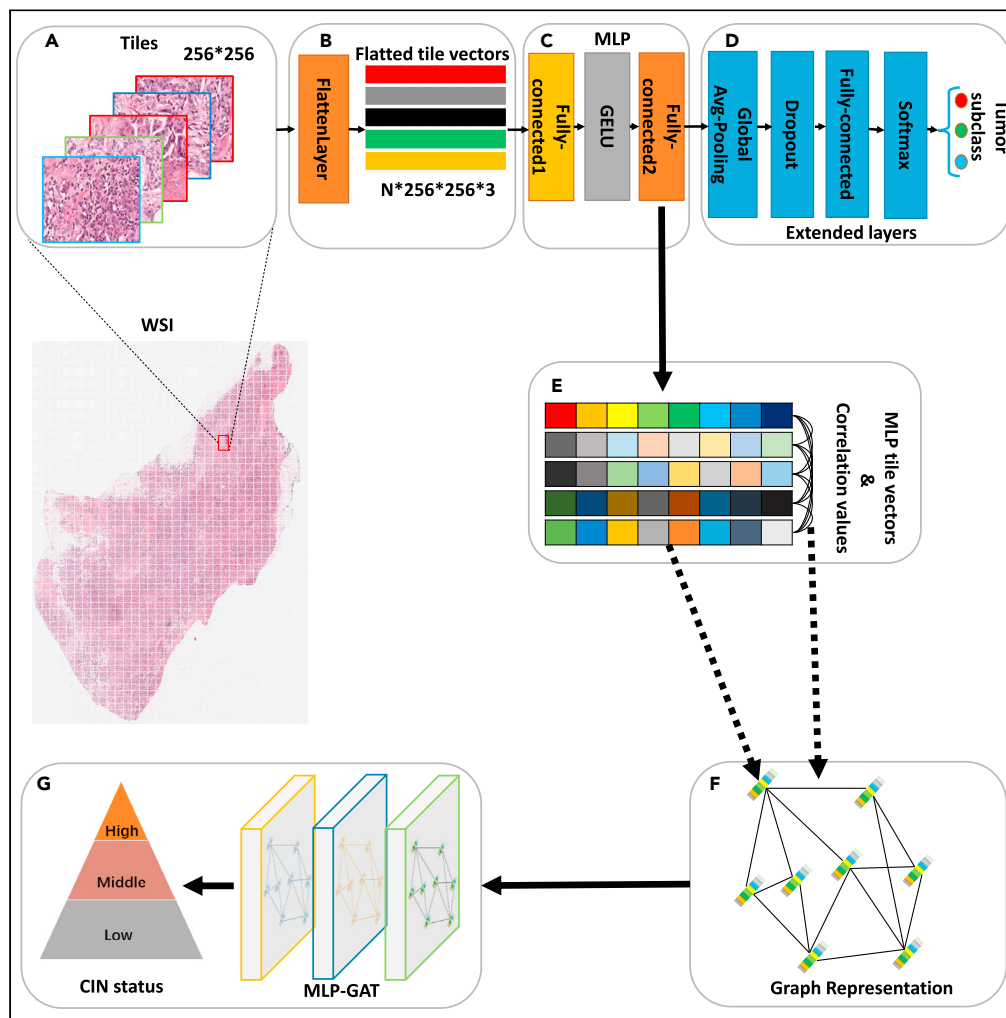


Article

A correlation graph attention network for classifying chromosomal instabilities from histopathology whole-slide images



Liangliang Liu, Ying Wang, Jing Chang, Pei Zhang, Shufeng Xiong, Hebing Liu

liangliu@henau.edu.cn (L.L.)
liuhebing@henau.edu.cn (H.L.)

Highlights

MLP-GAT was proposed to classify multi-level CINs from the WSI

The generalization performance of MLP-GAT was verified on the two cancer dataset

MLP-GAT provides guidance for studying CIN from the phenotype of cancer image



Article

A correlation graph attention network for classifying chromosomal instabilities from histopathology whole-slide images

Liangliang Liu,^{1,2,*} Ying Wang,¹ Jing Chang,¹ Pei Zhang,¹ Shufeng Xiong,¹ and Hebing Liu^{1,*}**SUMMARY**

The chromosome instability (CIN) is one of the hallmarks of cancer and is closely related to tumor metastasis. However, the sheer size and resolution of histopathology whole-slide images (WSIs) already challenges the capabilities of computational pathology. In this study, we propose a correlation graph attention network (MLP-GAT) that can construct graphs for classifying multi-type CINs from the WSIs of breast cancer. We construct a WSIs dataset of breast cancer from the Cancer Genome Atlas Breast Invasive Carcinoma (TCGA-BRCA). Extensive experiments show that MLP-GAT far outperforms accepted state-of-the-art methods and demonstrate the advantages of the constructed graph networks for analyzing WSI data. The visualization shows the difference among the tiles in a WSI. Furthermore, the generalization performance of the proposed method was verified on the stomach cancer. This study provides guidance for studying the relationship between CIN and cancer from the perspective of image phenotype.

INTRODUCTION**Clinical motivation**

The chromosome instability (CIN) is one of the hallmarks of cancer and is caused by continuous errors in chromosome segregation during mitosis. The CIN is the main driver of tumor evolution and can promote cancer metastasis by maintaining the autonomous response of tumor cells to cytoplasmic DNA.^{1,2} The status of the CIN not only determines intra-tumor heterogeneity and complexity but also has implications for carcinogens and therapeutics.^{1,3,4} For example, Smid et al.⁵ studied global patterns and the individual status of the CIN to determine the prognostic relevance between breast cancer subtypes. They suggested that an elevated CIN and consequent high aneuploidy burden is associated with poor breast cancer prognosis. Scribano et al.⁶ found a direct correlation between the CIN and the cancer treatment corner-stone paclitaxel in breast cancer patients, their findings supported the use of CIN as a predictive biomarker for the paclitaxel response. In addition, they suggested that increasing the CIN status improves the clinical utility of paclitaxel. Through a comparative analysis of 1,421 patients from 22 tumor types, Watkins et al.⁴ found that a continuous CIN status leads to widespread heterogeneity in somatic copy number, which affects the entire tumor-evolution process. Therefore, accurate classification and analysis of the CIN status is vital for cancer patients.

Although the CIN is clearly important for cancer diagnosis and treatment, due to the complexity of CIN detection, the professional testing equipment required for the biopsy and the complex analysis in clinical testing translate into a prohibitive consumption of human and financial resources. Histopathological whole-slide images (WSIs), which carry information of the biopsy, may be analyzed by computer algorithms to predict the CIN status. However, classifying the CIN status depends on the size of the WSI and the location of the target area. For example, Figure 1 shows a 728 MB WSI with dimensions of 97104 × 81862 × 3. Large images not only impose a severe computational burden but also pose serious challenges to the processing algorithms. As shown in the top row of Figure 1, the image contains significant background information that is irrelevant to the target tissue. The second row of Figure 1 shows that the three different magnifications present different details of the tissue. Figures 1A–1C show that weaker magnifications do not reveal the details of the tissue and that, as the magnification increases, the level of tissue detail also increases. However, as shown in Figure 1C, the scale of the clear detail image in the entire WSI is too small, which makes it difficult for the inspecting doctor or algorithm to find the target tissue within the sheer size

¹College of Information and Management Science, Henan Agricultural University, Zhengzhou, Henan 450046, P.R. China

²Lead contact

*Correspondence: liangliu@henau.edu.cn (L.L.), liuhebing@henau.edu.cn (H.L.)

<https://doi.org/10.1016/j.isci.2023.106874>



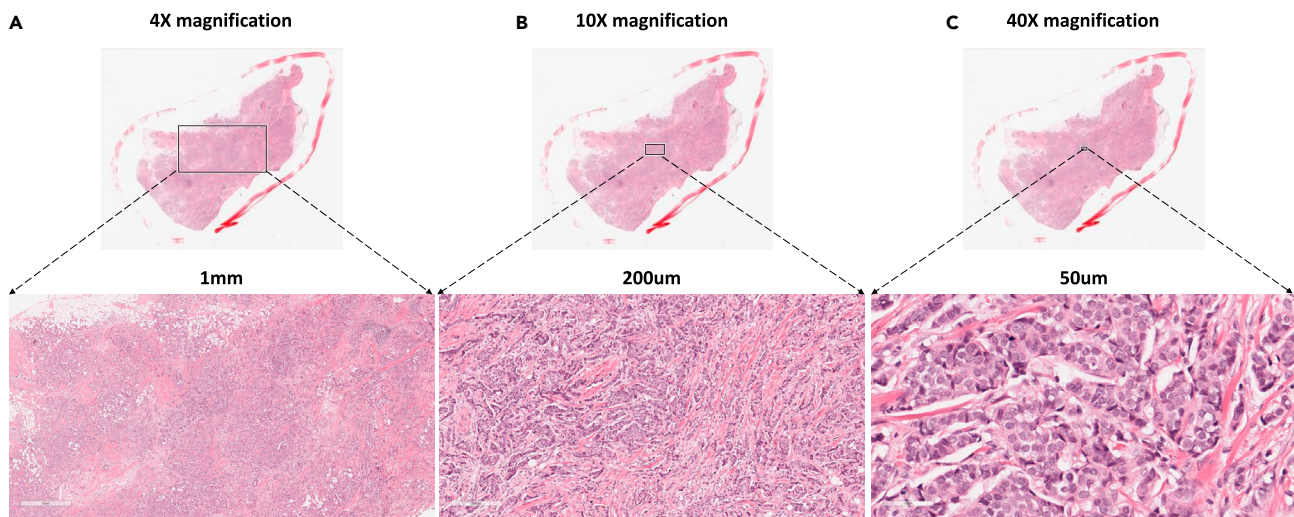


Figure 1. Illustration of a WSI with different magnifications

The top row shows the overall WSIs. The second row shows a part of the WSI with different magnifications. (A) A rectangular patch with 4× magnification, (B) a rectangular patch with 10× magnification, (C) a rectangular patch with 40× magnification.

and resolution of the WSIs. Therefore, how to balance the rich resolution and the WSI size is a topic worthy of study.

Although many previous studies have studied the CIN status, they mainly analyzed the relationship between the CIN and tumor via pathological detection or pathological experiments, which required significant manpower and material resources. To address this difficulty, an efficient algorithm is highly desirable to automatically classify the CIN status of tumors.

Related work

Significant progress has been made in the application of machine learning in various scenarios, including the prediction of CIN status in tumor patients.^{7–10} Most of these studies are based on gene expression data; the correlation between CIN status and gene expression is used to explore cancer treatment options and drug development. Singh et al.⁹ adopted the random forest technique to select features from gene expression profiles and developed a predictive model. They analyzed how CIN status affects dividing cells, thereby revealing potential mechanisms for an increase in CIN status in the late stage of papillary renal cell carcinoma. Shafat et al.⁸ established a suite of eleven single nucleotide polymorphism (SNP) array-based signatures of various forms of CINs and extracted a compendium of significantly associated genetic features by using a machine-learning method. Their study identified the signature of various forms of CIN that could provide potential new genomic targets for drug development for breast and ovarian cancer. In addition, Luxton et al.¹¹ used the XG boost model to analyze the relationship between CIN and genomic instability, thereby enabling risk assessment of the response of cancer patients to radiotherapy and adverse late health effects. Although machine learning-based methods have obtained relatively successful results for the association between the CIN and tumors, these studies were based on dozens of samples. In addition, the acquisition and cleaning of a large amount of genomics data requires significant financial and material resources. Given that WSIs of tumor tissue carriers exist in large numbers, the related research may be based on WSI data.¹²

Recently, deep learning-based methods have achieved a spectacular success in the biomedical image analysis tasks, including stroke segmentation, skin cancer diagnosis, retinal disease, brain tumor classification, and breast cancer.^{13–20} Many studies followed this trend and now use neural networks to learn patch-level feature information from WSI data. For example, Coudray et al.¹⁴ proposed a deep convolutional neural network (CNN; inception v3) on patch-wise features from WSIs to assess the stage, type, and subtype of lung tumors. The performance of inception v3 reached the level of the visual inspection by pathologists. Similarly, to reduce the workload of pathologists' visual inspection, Chen et al.²¹ used a neural network to classify and detect liver cancer gene mutation status on the patches of the hematoxylin and eosin (H&E) images. Fu et al.²² used a deep transfer learning network to classify multi-type cancer from H&E images. The experimental results show the significant potential of neural

networks in tumor histopathology research. To predict CIN status from patches in H&E images, Xu et al.²³ proposed a deep learning-based prediction model and found that tumor patients with higher CIN status translate into greater heterogeneity within the breast tumor. Although effective, these deep-learning methods use WSIs to assist the visual inspection by pathologists but still have to face the following challenges: (1) the large size of WSI. These methods usually adopted tissue segmentation and tile extraction to extract feature information from a WSI. For example, Bilal et al.²⁴ used ResNet34 method to predict the status of the CIN, key molecular pathways, and mutations by dividing the entire WSI into tiles of equal size. However, this strategy ignores the correlation between tiles, resulting in limited performance of the predictive model. (2) The presence of numerous background areas is a second challenge. Tumor-tissue blocks in WSIs always contain non-tumor tissue, which increases the difficulty for the predictive model to discover regions of interest (ROIs). (3) CNNs always learn over patch-wise features, which limit the ability of the model to capture global contextual information. Thus, without proper models, deep-learning methods that adopt a simple strategy of fragmenting large-scale WSIs have difficulty fully mining the features of WSI target tissues and the associated information between features.

Progress has recently been made in modeling the spatial information of the sheer size and resolution of WSIs using graph-based data and graph convolutional networks (GCNs).^{25–29} For example, Konda et al.²⁷ proposed a GCN model based on the manual selection features by the pathologist of the ROI of WSIs. The GCN was trained and tested on the graphs that were constructed from the nodes and edges of the tiles of a WSI. This GCN was applied to the resulting graph to aggregate the features for final classification on two image classification datasets. Similarly, based on manual selection ROIs, Zhenget al.²⁹ built a GCN-based diagnostic region attention network (DRA-Net). DRA-Net learns pathology knowledge from the image content within diagnosis paths and provides decision support for auxiliary diagnosis. In addition, Gao et al.³⁰ proposed a group quadratic GCN (GQ-GCN), which was executed on the graph that features WSIs obtained by a CNN. They used quadratic operations to enhance the GQ-GCN's feature selection capabilities. GCN-based research can enhance the prediction of the final target by modeling different ROI regions and the relationship between regions. At present, these methods construct maps by manually selecting ROIs. How to define the relationship between these ROIs is another problem requiring attention.

Contributions

In order to explore the method of CIN status prediction based on WSI, this study proposes a novel framework for classifying the CIN status from WSIs. Our method includes three parts: First, we fully exploit the sheer size and resolution of the WSIs and build a tile vector and correlation-based graph for each WSI. Second, an MLP-based GCN is proposed to classify the CIN status of breast cancer. Third, MLP-GAT is validated on the processed WSIs from public dataset TCGA-BRCA. In summary, our contributions are 3-fold.

- (1) We propose a correlation graph attention network (MLP-GAT) to classify multi-level CINs from the WSI of breast cancer.
- (2) We use the combination of the MLP and the cosine similarity (CS) methods construct the nodes and edges of the graph, which not only balance the spatial information and contextual information of a WSI but also bypass the WSI scale and resolution dilemma.
- (3) The generalization performance of the proposed method was verified on the stomach cancer. This study provides guidance for studying the relationship between CIN and cancer from the perspective of image phenotype.

The rest of the paper is organized as follows: Section "[methodology](#)" introduces the architecture of the proposed MLP-GAT. Section "[material and data preparation](#)" introduces the [material and data preparation](#). Section "[experiments and results](#)" shows empirical comparative and ablation experiments that investigate the contributions of the individual components of the proposed network. Finally, we discuss and summarize this study in Section "[discussion](#)".

Methodology

Overview of MLP-GAT

As shown in [Figure 2](#), with the input WSI data, MLP-GAT architecture consists of seven parts. [Figure 2A](#) shows the tiles with 256 × 256 pixels. None of the tiles overlap and all of the tiles of each object come from the ROI in a WSI. [Figure 2B](#) shows each tile is flattened into a tile vector with dimensions

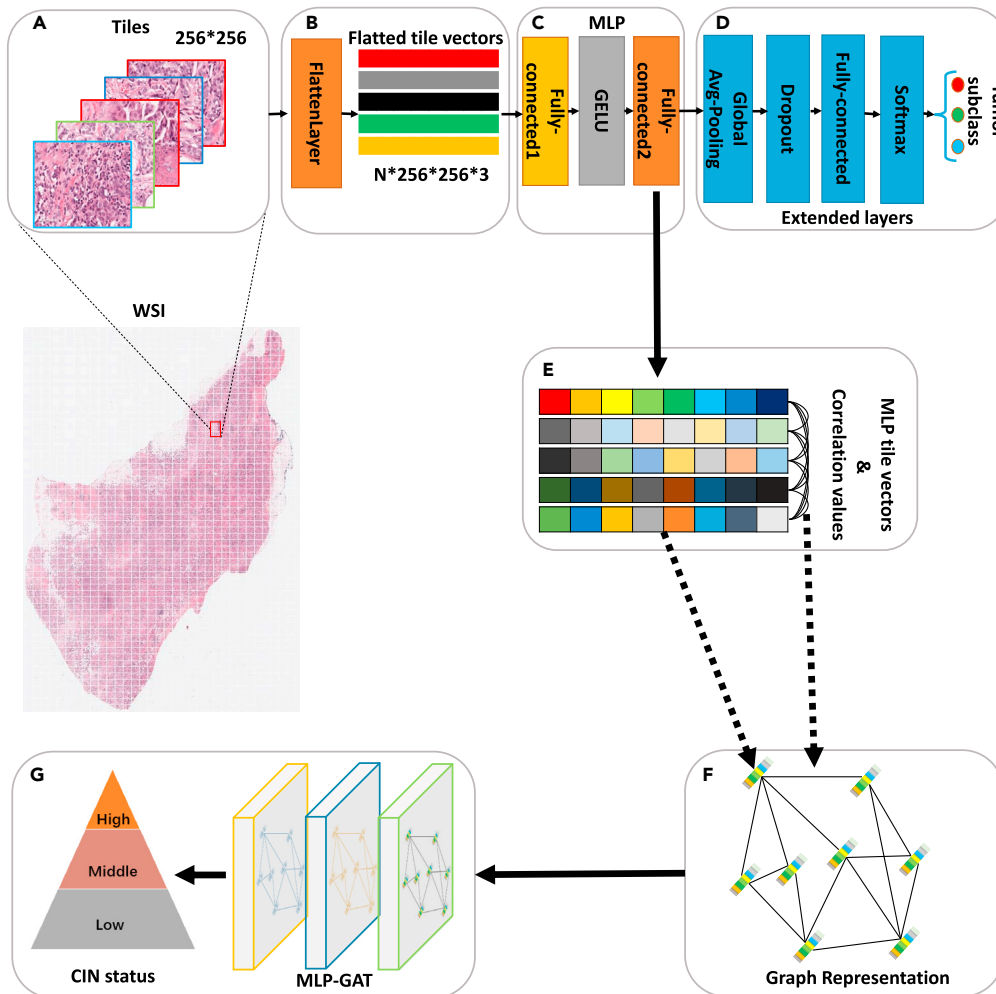


Figure 2. Overview of MLP-GAT architecture

The MLP generates tile feature vectors that are used to construct the graph vertices. The CS calculates the correlation coefficients between feature vectors, which are used to construct the graph edge weights. The MLP-GAT takes the graph as inputs and predicts the status of CIN. (A) shows the tiles with 256×256 pixels. (B) shows each tile is flattened into a tile vector with dimensions $256 \times 256 \times 3$. (C) shows the MLP network used to extract dimensionally reduced tile vectors. (D) shows the extended layers for tumor classification. (E) shows the determination of correlation values between any two tile vectors based on the CS method. (F) shows the representation of a constructed graph. (G) shows the proposed two-layer MLP-GAT and outputs the CIN status of the input WSI.

$256 \times 256 \times 3$. There are n tiles in a WSI and each tile has three channels. Figure 2C shows the MLP network used to extract dimensionally reduced tile vectors. Figure 2D shows the extended layers for tumor classification, which is used to fine-tune the MLP method. Figure 2E shows the determination of correlation values between any two tile vectors based on the CS method. Figure 2F shows the representation of a constructed graph. Finally, Figure 2G shows the proposed two-layer MLP-GAT and outputs the CIN status of the input WSI.

Processing whole-slide image

The sheer size and resolution of the WSI create severe challenges for image-analysis algorithms. In this study, we draw on the experience of other scholars to slice the WSI.^{24,27} As shown in Figure 2A, we divide the intermediate regions of each input WSI into 78 adjacent non-overlapping tiles and each tile with dimensions $256 \times 256 \times 3$. To facilitate the input of the fully connected layer in the MLP module, we use the Flatten Layer method³¹ to compress each tile of the same WSI into a one-dimensional vector, which is shown in Figure 2B.

Multi-layer perceptron

An MLP is a classical artificial neural network with forward structure, including input layer, output layer, and multiple hidden layers. The hidden layers of an MLP can learn the mapping relationship between the input layer and the output layer and realize transformation of feature dimensions. The layers of the MLP are fully connected.

The MLP model is a key part of the MLP-GAT model, each MLP consists of two fully connected layers and Gaussian error linear unit (GELU) nonlinearity.³² Let X be a one-dimensional vector input for MLP ($X \in \mathbb{R}^{m \times n}$) and m and n be the number and dimension of X , respectively. The output of the first fully connected layer H is then defined as:

$$H = XW_1 + b_1, \quad (\text{Equation 1})$$

where X denotes the input of the MLP. W_1 and b_1 are the weight and deviation of the first fully connected layer, respectively.

To increase the nonlinearity of the MLP model, we add a GELU nonlinearity activation function between the two fully connected layers. The GELU activation function comprehensively draws on the advantages and disadvantages of dropout, zoneout, and Relu. GELU is an input-dependent stochastic regularization method that takes the Bernoulli-distributed probability function $\text{Bernoulli}(\Phi(x))$ multiplied by the input neuron x . Specifically, we multiply x by a 0 or 1 mask (m). The Bernoulli-distributed probability function $\Phi(x)$ is defined as follows:

$$m \sim \text{Bernoulli}(\Phi(x)),$$

$$\text{where } \Phi(x) = P(K \leq x); K \sim N(0, 1), \quad (\text{Equation 2})$$

where K is a random variable, when $K \leq \Phi(x)$, $m = 0$, when $K \geq \Phi(x)$, $m = 1$. $\Phi(x)$ is the distribution function of the normal distribution, strictly increasing. Therefore, the smaller x is, the smaller $\Phi(x)$ is and the smaller the probability of $m = 1$; otherwise, the larger x is, the greater the probability of $m = 1$.

The GELU is defined as follows:

$$\text{GELU}(x) = xP(K \leq x) = x\Phi(x) = x \frac{1}{2} \left[1 + \text{erf} \left(\frac{x}{\sqrt{2}} \right) \right], \quad (\text{Equation 3})$$

where x is a neuron of the first fully connected layer output H .

The GELU scales a neuron x by probability. Specifically, if the output x_o of a neuron in the same layer is larger than the output of other neurons, the scaling factor of x_o is larger. Furthermore, $\text{GELU}(x)$ can be approximated as³²:

$$0.5x \left(1 + \tanh \left[\sqrt{\frac{2}{\pi}} (x + 0.044715x^3) \right] \right) \text{ or } x\sigma(1.702x). \quad (\text{Equation 4})$$

The output of an MLP can be defined as follows:

$$O = \hat{H}W_o + b_o, \quad (\text{Equation 5})$$

where \hat{H} , W_o , and b_o are the output of the GELU, the weight of the second fully connected layer, and the deviation of the second fully connected layer, respectively. Note that, when the dimension of O is less than X , MLP also plays a role in dimension reduction while extracting features.

To obtain the dimensionality-reduction features, a fully avg-pooling layer, a dropout layer, a fully connected layer, and a softmax activation function are successively added following the MLP model for supervised learning tumor classification (Figure 2D). This process is carried on the WSI-level. The tile size is 256×256 , the one-dimensional vector size of each flattened tile is 196608×1 ($256 \times 256 \times 3$), and we use the output vector of the second fully connected layer as the extracted feature of a tile.

Correlation analysis

In addition to node information, edge information is also an important part of a graph network. The edges reflect the strength of the association between two nodes. The interaction between nodes can be

accurately modeled if the correct association relationship is obtained. In our study, we use the CS method to measure the correlation relationship between two vectors in the extracted vector matrix. The obtained correlation coefficients represent the edge weights between any two nodes in a graph.

The CS method uses the cosine value of the angle between two vectors in vector space as a measure of the correlation coefficient between individuals.³³ The closer the correlation coefficient is to 1, the closer the included angle is to 0°; that is, the more similar the two vectors are. Let a and b be two vectors. The correlation coefficient can be defined as:

$$\cos(a, b) = \frac{\sum_{i=1}^n (x_i y_i)}{\sqrt{\sum_{i=1}^n (x_i)^2} \sqrt{\sum_{i=1}^n (y_i)^2}} = \frac{a \cdot b}{\|a\| \|b\|}, \quad (\text{Equation 6})$$

where n is the number unit in vector a or b, and x_i is the i-th unit in vector a (such as y_i in b).

Representation of constructed graph

The precise discrimination ability of the graph depends on the node values and edge weights. As shown in Figure 2B, we construct the nodes and edges of a graph using the MLP and CS methods, respectively. In the constructed graph, the connection relationship between any two nodes is determined by the correlation coefficient of the corresponding tile feature vectors in a WSI. Here, each WSI is mapped onto a vector matrix of size 78×128 , where each WSI keeps 78 slices. Thus, there are 78 nodes in a graph. Based on Section "correlation analysis", if the correlation coefficient between the vectors is greater than 0, the correlation coefficient is reserved as the edge weight between two nodes; otherwise, there is no connection between them.

MLP-GAT

Although GCNs can alleviate the challenges brought by WSI through tile images, GCNs treat all neighboring nodes equally during convolution operation and cannot assign different weights according to node importance. To solve this problem, we employ two measures: (1) We use the MLP method to construct the correlation between nodes in the graph (Sections "correlation analysis" and "representation of constructed graph" provides details). (2) A graph attention network (GAT) based on masked self-attention layers is proposed to assign the weight of different nodes in the adjacent node set.

The proposed MLP-GAT network is based on graph attention layers.³⁴ In an MLP-GAT network, we use the correlation coefficient as the attention coefficient between nodes v_i and v_j , the attention coefficient e_{ij} is defined as:

$$e_{ij} = \text{Attention}(v_i, v_j) = \cos(v_i, v_j). \quad (\text{Equation 7})$$

We directly aggregate the complete neighborhood $N(i)$ of a node v_i . We draw on the essence of masked graph attention³⁴ and only allow adjacent nodes to participate in the attention mechanism of the current node. The adjacent nodes are the first-order adjacent nodes (including the node itself). The attention score of node v_i is defined as:

$$a_{ij} \equiv \text{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in N(i)} \exp(e_{ik})}, \quad (\text{Equation 8})$$

where $\text{softmax}_j()$ is used to normalize the across nodes of j, which works as the attention coefficients for node v_i , which makes the attention coefficients between different nodes easy to compare. The $\exp()$ is an exponential function.

These attention scores can be used to define the propagation rule for the GAT layer. In the training process, the attention scores a_{ij} are used to weight the different contributions v_j for the update of v_i based on the coefficient value (correlation coefficient). To stabilize the self-attention learning process, we use a multi-head attention mechanism, which is defined as:

$$x'_i = \left\| \frac{P}{\rho} = 1 \right. \sigma \left(\sum_{j \in N(i)} a_{ij}^p \Theta_j^p \right), \quad (\text{Equation 9})$$

where P refers to the number of used heads, Θ_j^P represents a learned linear transformation of node j . Specifically, the P independent attention mechanisms are transformed according to Equation 9 to obtain updated node output features, and then the P features are concatenated and transferred to the next GAT layer. Only the last layer performs an averaging of the different received representations.³⁴

In terms of GAT optimization, we draw on the experience of Chen et al.,³⁵ MLP-GAT consists of three graph attention hidden layers. We perform the averaging operation on all node features in the last layer and make the classifying topics using the softmax activation function.

Material and data preparation

Dataset

The WSIs along with clinical and genomic data were downloaded from the TCGA-BRCA (<https://portal.gdc.cancer.gov/>). This dataset contains clinical, genetic, and pathological image data of patients, allowing data mining analysis by correlating cancer phenotype with genetic information. There are 1,098 patients covering five carcinoma types, including 778 cases of ductal carcinoma (DC), 201 cases of lobular carcinoma, 16 cases of mucinous carcinoma, 8 cases of papillary carcinoma, and the remaining cases are mixed carcinoma. The formalin-fixed paraffin-embedded diagnostic H&E-stained histopathology WSIs from primary tumor sites are the subject of our study. After removing WSIs that lack magnification information, and/or WSIs with artifacts, including tissue folding, air bubbles, and out-of-focus regions, we get 896 cases with 896 WSIs, including 680 cases of ductal carcinoma, 200 cases of lobular carcinoma, and 16 cases of mucinous carcinoma.

We follow the method of Salas et al. and adopt segments with log-transformed mean copy number values as the basis for calculating CIN status.³⁶ Specifically, we use whole exome sequencing-based copy number variation data to calculate the fractional genome altered value of the same cases. By examining the fractional genome altered distribution for all 896 cases, we label cases with fractional genome altered values less than 0.3 as low CIN; those with fractional genome altered value greater than or equal to 0.3 as high CIN and the remainder as moderate CIN.

Image preprocessing

The sheer size and resolution of the WSIs limit the performance of the algorithms. In addition, the white background and noise in WSI would be a challenge. Therefore, we use a single overall ROI with dimensions of 20000 × 20000 pixels at 40× magnification to represent a WSI. For each WSI, the white background accounts for almost half of the WSI area, but has nothing to do with pathological tissue. Therefore, the chosen ROI avoids the white background as much as possible. We define the intermediate region of a WSI by obtaining the WSI center coordinate. We obtain the center coordinate according to the size of WSI, and then select the 20000 × 20000 pixels as the ROI area with the center coordinate. We use a sliding window with dimensions of 256 × 256 pixels to slide in the ROI area and extract 78 tiles from each WSI without overlap. In addition, we normalize each tile using the Python library (HistomicsTK) to convert the color features to the desired annotations to reduce the batch effect between different WSI.³⁷ Overall, we get 78 tiles from each WSI, the labels of these tiles are consistent with the WSI from which they came.

Experiments and results

Experimental setup

There are three steps to train the entire pipeline. We first train the MLP and extended layers to produce tile vectors in the supervised manner. We then use the CS method to generate correlation coefficients. Finally, we construct graphs that are used to train and test the MLP-GAT. All the experiments are implemented in Pytorch on a papillary carcinoma with an NVIDIA GeForce Titan X Pascal CUDA GPU processor. There are four metrics to evaluate the performance of the proposed network, including accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the receiver operating characteristic curve (AUC).

Tile vector and correlation coefficient learning

As shown in Figures 2C and 2D, the MLP is used to extract a tile vector and extended layers are used to predict the subtype of tumors in the supervised manner. We extract tile vectors for 896 cases. At this stage, 896 cases are divided training set and testing set (8:2), the training set is used to train and fine-tuning MLP and the extended layers, the testing set is used to verify the performance of this proposed method. As

Table 1. Classification results of the six comparative methods on multi-type CINs

Method	Image type	ACC (%)	SEN (%)	SPE (%)	AUC (%)
AlexNet ⁴²	10× magnification	52.74(±10.35)	54.80(±8.47)	53.97(±10.48)	58.74(±11.06)
AlexNet ⁴²	40× magnification	54.01(±7.87)	55.16(±9.73)	55.07(±9.42)	59.37(±6.57)
VGG16 ⁴³	10× magnification	52.98(±6.45)	56.83(±9.04)	51.98(±11.08)	58.96(±7.89)
VGG16 ⁴³	40× magnification	58.87(±9.16)	59.72(±10.35)	59.30(±9.87)	61.04(±10.18)
GoogLeNet ⁴⁴	10× magnification	58.07(±12.17)	57.68(±6.79)	59.67(±11.52)	60.78(±9.48)
GoogLeNet ⁴⁴	40× magnification	71.51(±8.42)	69.78(±11.56)	71.98(±13.18)	73.53(±12.60)
ResNet ⁴⁵	10× magnification	65.09(±11.71)	67.13(±6.78)	66.54(±9.78)	68.51(±7.83)
ResNet ⁴⁵	40× magnification	72.61(±8.21)	72.86(±9.27)	73.01(±12.08)	74.56(±8.43)
Transfer Learning ⁴⁶	10× magnification	68.19(±4.89)	69.76(±10.05)	67.18(±9.46)	71.03(±13.42)
Transfer Learning ⁴⁶	40× magnification	74.25(±12.17)	75.43(±9.14)	73.88(±10.04)	75.81(±6.79)
ResNet34 ²⁴	10× magnification	67.41(±7.41)	69.02(±8.37)	66.80(±8.91)	69.26(±11.43)
ResNet34 ²⁴	40× magnification	73.37(±8.72)	75.01(±10.54)	72.59(±11.41)	75.42(±8.84)
GCN-Hash ²⁹	10× magnification	71.04(±5.21)	72.69(±5.10)	73.98(±4.69)	73.86(±10.42)
GCN-Hash ²⁹	40× magnification	77.67(±7.12)	76.87(±9.75)	76.09(±10.11)	79.67(±9.68)
MLP-GAT	10× magnification	74.97(±11.73)	75.08(±7.90)	75.21(±7.39)	79.16(±10.37)
MLP-GAT	40× magnification	79.02(±9.73)	80.04(±8.74)	78.49(±8.72)	82.06(±11.68)

Values correspond to the mean(standard deviation). The best results appear in bold.

introduced in Section "multi-layer perceptron", we use the flattened tile vector of dimensions 196608×1 as the input for the MLP, we use the output vector of the second fully connected layer of the MLP as the extracted feature of the input tile, with dimensions 128×1 . Finally, each WSI can be mapped into a vector matrix of size 78×128 . In this study, we use these vectors to construct nodes in the graph network. In addition, we use the CS method to measure the correlation coefficient between two tile vectors, as introduced in Section "correlation analysis". We use these tile vectors and correlation coefficients to construct the graph on the WSI-level.

Comparison with state-of-the-art methods

Lots of studies have shown that CNNs are excellent for analyzing medical images.^{38–41} Most of these methods used convolutions to extract the contextual semantic information for pixels on the image level (as opposed to the tile level). Therefore, we conduct extensive 2-fold experiments on multi-type CIN classification: (1) we use WSIs as input for the comparison methods. To avoid excessively large WSIs, the comparison methods are conducted on the processed tiles of 10× magnified WSIs with dimensions 256×256 and tiles of 40× magnified WSIs with dimensions 256×256 . (2) We conduct MLP-GAT on the constructed graphs obtained from the 10X and 40X magnification WSIs (Section "image preprocessing" provides details), respectively. The comparison methods are derived from code published by the author or from reconstructions of the models in their papers. Furthermore, we adjust the comparison methods to the size of the tile. To ensure fairness, these experiments were conducted in the training set and testing data with a ratio of 8:2.

We use two sizes of tiles as inputs for these methods. Table 1 lists the comparative results. The same method based on 40× magnified tiles produce a better classification than that based on 10× magnified tiles, mainly because the 40× magnified tiles offer much more detail of the pathological tissues than the 10× magnified tiles (as shown in Figure 1). In addition, except for MLPGAT, the GCN-based method and transfer learning-based method achieve the top two classification results (accuracy = 77.67% vs. accuracy = 74.25%). The accuracy of the proposed MLP-GAT improves by 1.35% on GCN-Hash method. The transfer learning method⁴⁶ is mainly limited to mining the relationship among tiles. While GCN-Hash method²⁹ can mine the relationship among tiles, but it does not pay attention to the weight relationship among these tiles. Other metrics of MLP-GAT are also the best of all methods, mainly because the other methods directly use the pixel context semantic information in the tile and make predictions but ignore the correlation between tiles. However, the non-overlap tiles in a WSI inevitably lead the pathological structure of the edge to be torn apart, which affects the extraction of semantic information from the edge organization context. In the MLP-GAT method, we pay attention to the relationship between different tiles.

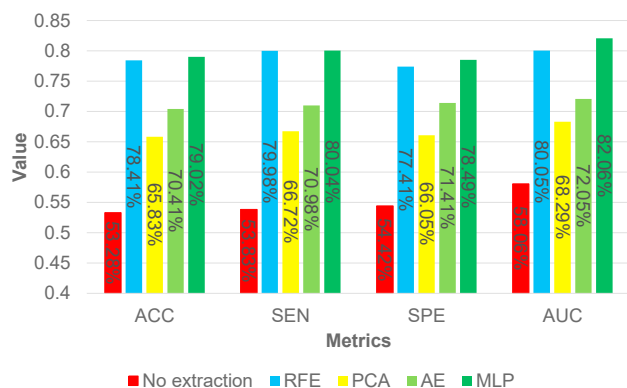


Figure 3. Histograms of different feature vector extraction strategies on MLP-GAT

“No extraction” denotes vectorization of all pixels in the tile as feature vectors.

We use the CS method to convert this relationship into the weight of edges in the graph, which makes up for this relationship from another angle. Experimental results indicate that this strategy is effective. Therefore, the MLP-GAT method produces the best classification results.

Feature vector extraction strategy

The MLP is the key part of our framework and serves to extract feature vectors from tiles. We construct the nodes of graphs using the feature vectors of the dimension 128×1 . We train and evaluate the GAT by using the construction graphs for 128 features. To verify the role of the MLP in the framework, we compare it with several other feature extraction methods, including recursive feature elimination (RFE), principal component analysis (PCA), and auto-encoder (AE). We use a ridge classifier to perform RFE with a fixed number of features.⁴⁷ PCA is commonly used to reduce dimensions and uses singular value decomposition to project the data into a lower-dimensional space.⁴⁸ AE consists of unsupervised neural networks that aim to learn a lower-dimensional representation from an input data.⁴⁹ We use RFE, PCA, AE, and MLP to extract feature vectors of dimensions 128×1 from the tiles of dimensions 256×256 . We train and evaluate the GAT on graphs constructed from these feature vectors.

The results are shown in [Figure 3](#): all metrics for the four feature vector extraction methods are under the 10-fold cross-validation strategy. “No extraction” means that there is no feature extraction; we only use the flatten tile vector as GAT input. Overall, the methods without feature extraction (“No extraction”) produce the worst classification result. This is mainly due to the high sparsity and noise of tiles. In addition, the classification performance of PCA-based GAT is also very poor, which is mainly because PCA is not adapted to the problem that the dimensions of the feature vector are much greater than the number of samples. Compared with PCA, the accuracy and AUC of GAT based on the unsupervised AE method are improved almost by 5% and 4%, respectively. Supervised RFE- and MLP-based feature extraction methods help GATs achieve the best classification results, which is because the unsupervised approach does not pay attention to the feature of the target. MLP-based GAT methods are even better for the feature vector of dimension 128×1 . This may be because MLP inherently possess feature transformation capabilities, which has been demonstrated to be of significant value in vision tasks.^{50–52}

Influence of size of feature vectors

We combined the MLP and extended layers to extract feature vectors from tiles. The best feature vectors are extracted based on the ability of trained fully connected layers to predict the tumor type of breast cancer. The results show that MLPs achieve the best performance, with an accuracy = 86.73% and AUC = 89.16%. We extract a feature vector with dimensions of 128×1 for a tile from the second fully connected layer in the MLP. To verify how different feature vectors affect on the MLP-GAT model, we extract feature vectors of five different sizes (64, 128, 256, 512, and 1,024) by using the MLP.

The results of the comparison appear in [Table 2](#). On graph networks constructed based on feature vectors of different sizes, the accuracies of MLP-GATs all exceed 70%, except for MLP-GAT(64). This is limited by the

Table 2. Influence of different feature vectors on the multi-type CINs classification

Method	ACC (%)	SEN (%)	SPE (%)	AUC (%)	Training time (hour)
MLP-GAT(64)	67.08(±4.95)	68.76(±9.18)	67.83(±11.51)	69.64(±8.42)	23.5
MLP-GAT(128)	79.02(±9.73)	80.04(±8.74)	78.49(±8.72)	82.06(±11.68)	26.7
MLP-GAT(256)	79.35(±12.02)	78.12(±8.43)	78.62(±9.40)	81.98(±11.02)	31.4
MLP-GAT(512)	77.08(±9.43)	77.86(±8.45)	76.51(±6.73)	80.01(±9.16)	37.1
MLP-GAT(1024)	75.36(±10.64)	76.18(±5.72)	75.94(±6.80)	79.69(±8.49)	42.3

Values correspond to the mean(standard deviation). The best results appear in bold.

minimal feature information in 64×1 vectors. MLP-GAT(512) and MLP-GAT(1,024) obtain much more information from 512×1 and 1024×1 vectors, respectively. However, they both may also contain noise features, resulting in poor model performance and excessive training time. The four main evaluation metrics of MLP-GAT(128) and MLP-GAT(256) are the best, but MLP-GAT(256) requires almost 5 h more training time than that of MLP-GAT(128). In this study, we use a feature vector of dimensions 128×1 to construct the node of a graph.

Correlation coefficient calculative strategy

In this study, we use the MLP to obtain the feature vectors from tiles and use the CS method to measure the similarity between two feature vectors. Feature vectors represent the nodes in a graph. The similarity coefficients represent the edge weights between any two nodes in the graph. This is a contribution of the MLP-GAT. To demonstrate the effectiveness of feature vectorization and similarity coefficient methods on MLP-GAT, we undertook several experiments on the similarity coefficient between feature vectors. The similarity coefficient methods include "no similarity", euclidean distance (ED), hamming distance (HD), structural similarity (SSIM), mutual information (MI), dHash, and CS. "No similarity" means that two feature vectors share similarities between them. ED measures the distance occupied by an individual. HD gives the number of different bits corresponding to two vectors of the same length. SSIM measures the similarity between pictures and can also be used to judge the quality of compressed pictures. MI gives the similarity between two images by calculating their mutual information. dHash is called the "different hashing algorithm." When hashing images, dHash obtains the final hashing sequence by comparing the size of the left and right pixels.

Figure 4 shows the comparisons. The first observation is the very poor performance of the "no similarity" strategy. This strategy-constructed graph only focus on the contextual semantic information within the tiles while ignoring the hidden association information between the tiles. However, several other similarity-calculation methods (ED, MI, HD, SSIM, dHash, and CS) consider the similarity information between feature vectors. For example, ED and HD mainly focus on distance similarity, SSIM mainly focuses on structural similarity, MI is mainly used to measure of random variables, and dHash mainly focuses on the similarity of image feature brightness and contrast. Although CS is more sensitive to vectory image information, it calculates the cosine distance between vectors to characterize the similarity of two vectors and is beneficial to the MLP-GAT model. Therefore, CS is chosen as the method of similarity coefficient in our experiment.

DISCUSSION

Heterogeneity among tiles

In this study, we extracted tile feature information based on the classification of breast cancer tumors using MLP and extended layers. The pathological histology corresponding to tiles from the same WSI is related to the subtype of breast cancer. Different pathological tissues and subtypes have different degrees of compactness. Therefore, we take the WSI of a DC sample as an example and selected 64 tiles for subtype prediction probability analysis to verify the heterogeneity among tiles in the same WSI.

Figure 5 shows the visualization results. Figure 5A is a macro-image that shows the location of the ROI in a WSI. Figure 5B is a micro-image that shows the detail of the predicted probability of each tile. For the same breast cancer subtype, the predicted probabilities of different tiles are quite different. This means that such probabilistic differences also lead to differences in extracted features, so nodes that use these extracted features as component graphs can reflect the organizational information in the original WSI. Moreover, in Figure 5B, the predicted probabilities between adjacent tiles also differ. In this study, we use the CS method to quantify this difference and to map it to the component graph.

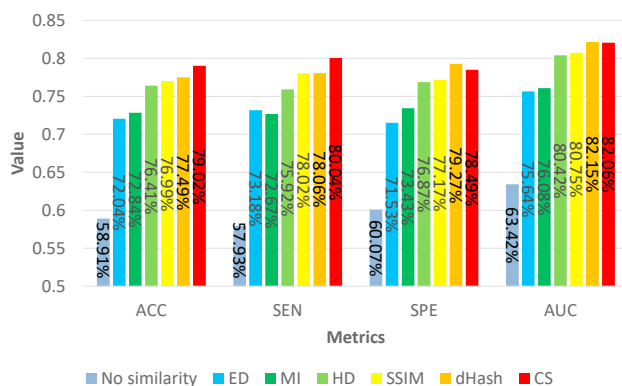


Figure 4. Histograms of different similarity methods on MLP-GAT
“No similarity” denotes there are no similarities between any two feature vectors.

Influence of tile position

In clinical diagnosis, formalin-fixed paraffin-embedded diagnostic H&E-stained histopathology slides form the key basis used by doctors to analyze the pathological tissue of breast tumors. We use non-overlapping tiles to study CINs in breast cancer. In [Figure 6](#), we select six representative tiles to analysis at 40× magnification to verify how tile position affects model performance. These are six tiles that come from the same WSI of DC. After identification by pathological experts, [Figure 6A](#) shows a tile of normal cell tissue. [Figures 6B](#) and [6C](#) show tiles for which experts cannot determine whether they contain tumor cells, and [Figures 6D–6F](#) show tiles that contain cancer cells.

As shown by the yellow square in [Figure 6A](#), the structural display of cells is affected by the problem of staining for WSIs. The existence of such tiles affects the accuracy of features extracted by MLPs. In addition, the morphology of the cells in the yellow boxes indicated by the red arrows in [Figures 6B](#) and [6C](#) are destroyed when the tiles are sliced, which causes the tumor classification model to divide the tiles into a normal group of tiles. We invited the pathological-imaging experts were also ambivalent about such tiles. In addition, experts confirmed that the slices in [Figures 6D–6F](#) contain cancerous cells, but the fraction of cancerous cells in these three tiles is too small, so the feature extraction model focuses on the pixels in the background area. Finally, MLP-GAT ignores the pixel information of the cancer cells, thus causing the model to misclassify the three tiles. As part of the constructed graph, the presence of these types of tile misleads the classification by MLP-GAT.

Balance computations and overfitting

In this study, the MLP is the basis of feature extraction and graph construction. An MLP requires numerous parameters, which means that it requires significant computing power and data. Given insufficient data, an MLP easily falls into overfitting problem. Therefore, MLPs usually perform well with small models but then seriously overfitting problem as the model size increases. To avoid this problem, we conduct several experiments with different tile sizes before constructing the graph. We adopt 128 × 128, 256 × 256, 512 × 512, and 1024 × 1024 as the tile size. The experimental process is shown in [Figures 2A–2D](#), where we adopt tumor classification accuracy as the main criterion for selecting tiles of different sizes. The experimental results appear in [Table 3](#).

The results given in [Table 3](#) show that the tile of size 128 × 128 produces the best tumor classifications. The larger the tile size, the lower the performance of the MLP. This is because the number of tiles extracted from a given WSI correlates negatively with the tile size. The characteristics of the MLP structure require a large amount of data to avoid overfitting problems. The loss curves of training and validation of the MLP method for these four tile sizes validate this conjecture (see [Figure 7](#)). In the training phase, the loss curves of the four tiles gradually decrease with increasing epoch, whereas in the verification phase, the tile loss curves for 128_128 and 256 × 256 gradually converge, but the loss curves for 512 × 512 and 1024 × 1024 suffer from overfitting.

With the proposed method, we select 256 × 256 as the tile size instead of 128 × 128 because (1) the small-size tiles easily lead to unbalanced samples. The samples that contained and did not contain cancer cells

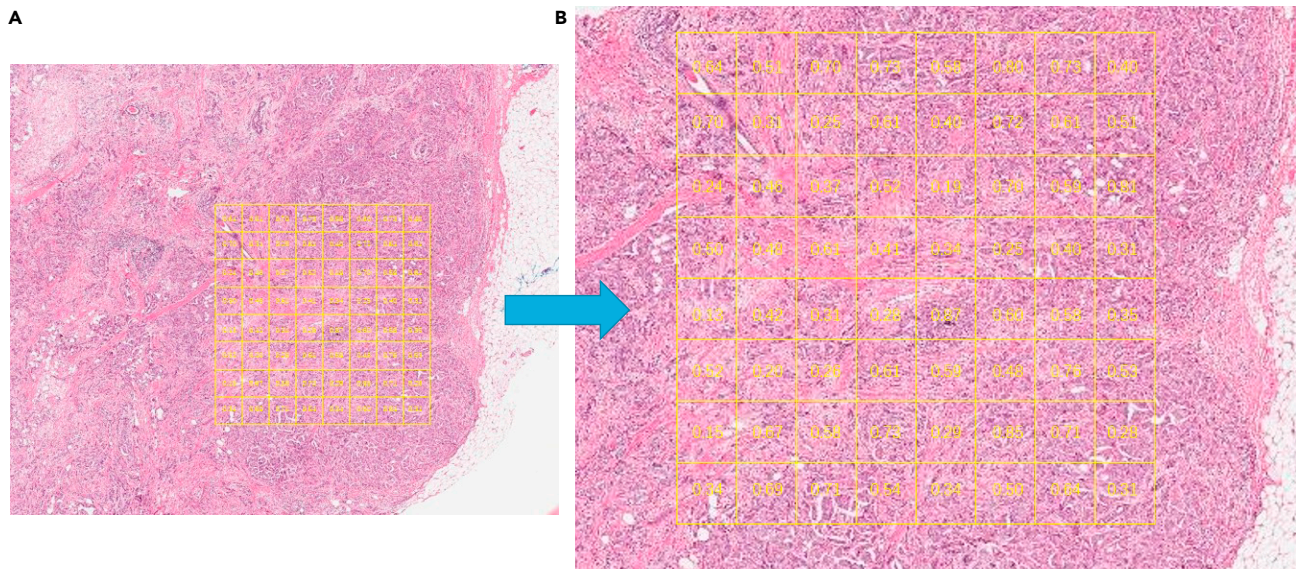


Figure 5. Intra-tumor heterogeneity by patch predictions

(A) shows the location of ROI area in a WSI.

(B) Shows the detail of the predicted probability of each tile.

differ too much. (2) In addition, to represent the ROI of the same size, the graph constructed from tiles of size 128×128 is much larger, and the amount of computation required is almost three times (3×26.7 h) greater than the graph constructed based on tiles of size 256×256 . In summary, to balance the calculation load with overfitting, we use 256×256 as the tile size.

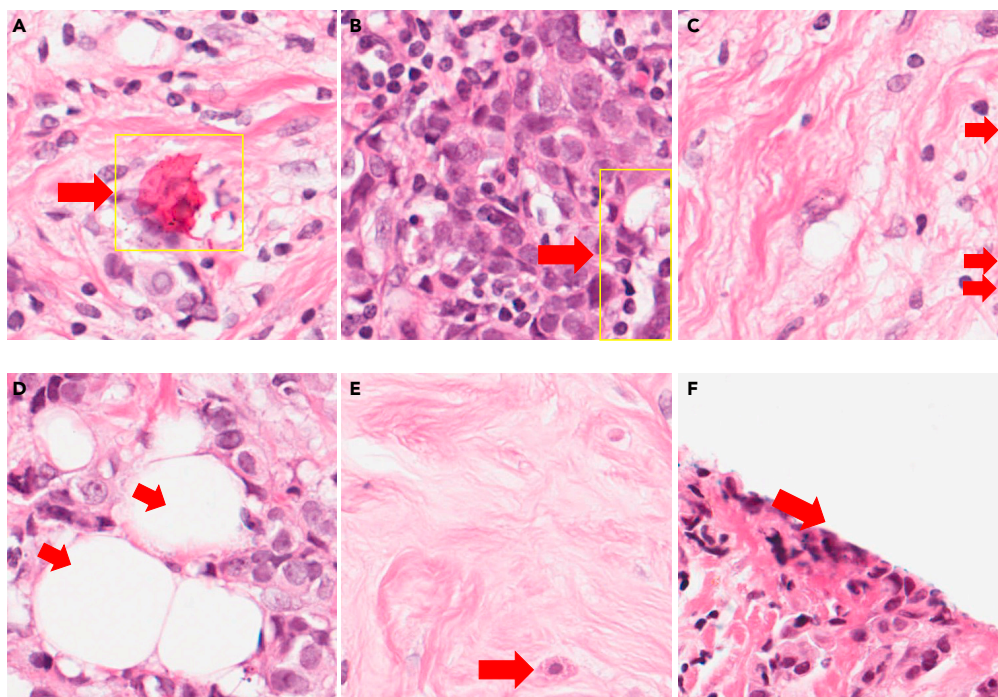


Figure 6. Visualization of six special tiles

(A) shows a tile of normal cell tissue.

(B and C) show tiles for which experts cannot determine whether they contain tumor cells.

(D–F) show tiles that contain cancer cells.

Table 3. Influence of tile size on MLP of the multi-type tumor classification

Tile size	ACC (%)	SEN (%)	SPE (%)	AUC (%)
128 × 128	90.13(±6.75)	89.84(±10.17)	91.06(±6.59)	91.43(±9.04)
256 × 256	86.73(±7.19)	87.08(±4.86)	85.97(±9.14)	89.16(±6.82)
512 × 512	78.06(±8.43)	76.31(±9.47)	77.28(±7.42)	79.54(±12.07)
1024 × 1024	75.39(±10.27)	76.08(±7.64)	74.59(±11.23)	77.15(±7.64)

Values correspond to the mean(standard deviation). The best results appear in bold.

External validation

To validate the performance of the proposed MLP-GAT, we extended the experiment on stomach cancer. We downloaded 102 cases from TCGA-STAD (<https://portal.gdc.cancer.gov/>). There are 65 cases with high CIN and 37 cases with low CIN. We processed the downloaded images according to the data processing process in Section "image preprocessing". We used the MLP method to construct the correlation between nodes in the graph (Sections "correlation analysis" and "representation of constructed graph"). As shown in the following Table 4, our proposed MLP-GAT has the advantage of 1.03% on accuracy and other evaluation indexes of MLP-GAT are also the best. It denotes that the effectiveness of the proposed MLP-GAT. At the same time, it cannot be ignored that the performance of transfer learning, ResNet34, GCN-Hash, and MLP-GAT is relatively close. This is mainly due to the limitation of sample size and data imbalance in dataset TCGA-STAD. Our future work is to increase the sample size and find new validation datasets. Furthermore, compared to the performance of our proposed MLP-GAT on TCAG-BRAC and TCGA-STAD, there are differences in the metrics, which implies that the proposed model may perform differently on various types of cancer.

Conclusion and future perspectives

We proposed the MLP-GAT network which can classify multi-type CINs from breast cancer WSIs. We realize multi-classification based on the constructed graphs and use a GAT network to implement classification. Extensive experiments confirm that the MLP-GAT delivers the state-of-the-art performance and that the feature extraction and similarity computation methods are effective. We further analyze the relationship between breast cancer subtypes and the CIN. We found the difference between tiles in WSI of breast cancer, which provides new ideas for the study of this disease. In future work, we plan to further improve the architecture of the network to address current limitations and we will try to apply this method to other cancer research.

Limitation of the study

There are some limitations in our study. First, we select the 20000 × 20000 pixels as the ROI area with the center coordinate and the sliding window obtains 78 tiles from the ROI area, which is challenging to cover

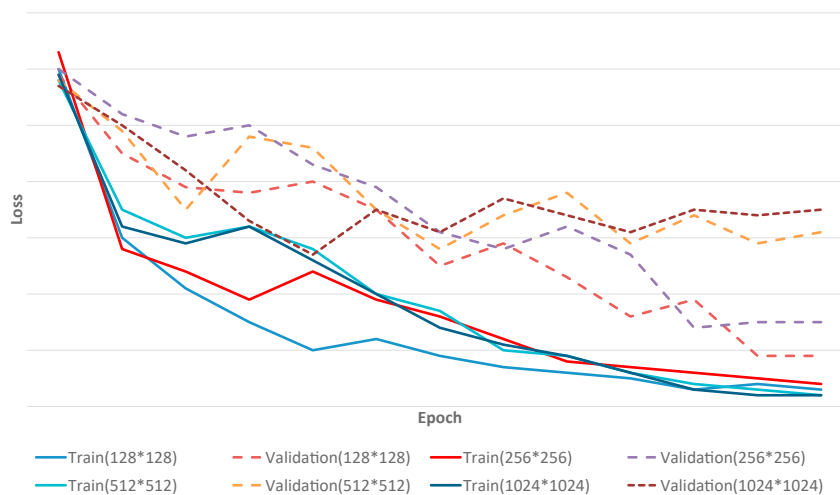


Figure 7. Loss curves of training and validation on MLPs for four tile sizes

Table 4. External validation on stomach cancer from TCGA-STAD

Method	ACC (%)	SEN (%)	SPE (%)	AUC (%)
AlexNet	50.47	50.82	50.70	51.06
VGG16	52.35	52.76	51.89	52.98
GoogLeNet	52.99	53.04	52.76	53.58
ResNet	61.08	61.42	60.93	62.63
Transfer Learning	66.51	66.89	66.74	68.89
ResNet34	66.49	67.08	67.26	68.17
GCN-Hash	67.09	67.82	67.54	68.97
MLP-GAT	68.12	68.76	68.48	70.34

the scope of all WSI and results in information loss. In addition, as shown in [Figure 6](#), the performance of MLP-GAT is affected by tile position. How to better mine the spatial relationships and spatial location of tiles is also the direction of our model improvement. Finally, due to not having collected the “true” external validation dataset related to breast cancer to validate the proposed MLP-GAT, we validated the proposed model using TCGA-STAD data. However, TCGA-BRCA and TCGA-STAD are two different types of cancer. The experimental results indicate that the proposed model performs differently on these two types of cancer, which is also one of the limitations of our model.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [RESOURCE AVAILABILITY](#)
 - Lead contact
 - Materials availability
 - Data and code availability
- [EXPERIMENTAL MODEL AND SUBJECT DETAILS](#)
 - Participants
- [METHOD DETAILS](#)
 - Overview of MLP-GAT
 - Processing whole slide image
 - Multi-layer perceptron
 - Correlation analysis
 - Representation of constructed graph
 - MLP-GAT
 - Experimental setup
 - Tile vector and correlation coefficient learning
- [QUANTIFICATION AND STATISTICAL ANALYSIS](#)
 - Significance results

ACKNOWLEDGMENTS

This work was funded by the Henan Provincial Key Research and Promotion Projects (No. 222102310085) and the Key R&D projects during the 14th Five Year Plan period (No.2022YFD1400302).

AUTHOR CONTRIBUTIONS

L. L., S. X., and H. L. conceived and designed the experiments. L. L. and J. C. performed the experiments. Y. W. and P. Z. analyzed and interpreted the data. L. L. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no conflict of interest.

Received: November 15, 2022

Revised: February 23, 2023

Accepted: May 10, 2023

Published: May 18, 2023

REFERENCES

- Bakhom, S.F., and Cantley, L.C. (2018). The multifaceted role of chromosomal instability in cancer and its microenvironment. *Cell* 174, 1347–1360.
- Bakhom, S.F., Ngo, B., Laughney, A.M., Cavallo, J.-A., Murphy, C.J., Ly, P., Shah, P., Sriram, R.K., Watkins, T.B.K., Taunk, N.K., et al. (2018). Chromosomal instability drives metastasis through a cytosolic dna response. *Nature* 553, 467–472.
- Guo, Y.A., Chang, M.M., Huang, W., Ooi, W.F., Xing, M., Tan, P., and Skanderup, A.J. (2018). Mutation hotspots at ctf binding sites coupled to chromosomal instability in gastrointestinal cancers. *Nat. Commun.* 9, 1520–1614.
- Watkins, T.B.K., Lim, E.L., Petkovic, M., Elizalde, S., Birkbak, N.J., Wilson, G.A., Moore, D.A., Grönroos, E., Rowan, A., Dewhurst, S.M., et al. (2020). Pervasive chromosomal instability and karyotype order in tumour evolution. *Nature* 587, 126–132.
- Smid, M., Hoes, M., Sieuwerts, A.M., Sleijfer, S., Zhang, Y., Wang, Y., Foekens, J.A., and Martens, J.W.M. (2011). Patterns and incidence of chromosomal instability and their prognostic relevance in breast cancer subtypes. *Breast Cancer Res. Treat.* 128, 23–30.
- Zhou, A.S., Zasadil, L.M., Molini, R., Fitzgerald, J., Lager, A.M., Laffin, J.J., Correia-Staudt, K., Wisinski, K.B., Tevaarwerk, A.J., and O'Regan, R. (2021). Chromosomal instability sensitizes patient breast tumors to multipolar divisions induced by paclitaxel. *Sci. Transl. Med.* 13, eabd4811.
- Luxton, J.J., McKenna, M.J., Lewis, A.M., Taylor, L.E., Jhavar, S.G., Swanson, G.P., and Bailey, S.M. (2021b). Telomere length dynamics and chromosomal instability for predicting individual radiosensitivity and risk via machine learning. *J. Personalized Med.* 11, 188.
- Shafat, M.S., Rufaie, E.S., and Watkins, J. (2021). A survey of cancer genome signatures identifies genes connected to distinct chromosomal instability phenotypes. *Pharmacogenomics J.* 21, 390–401.
- Singh, N.P., Bapi, R.S., and Vinod, P.K. (2018). Machine learning models to predict the progression from early to late stages of papillary renal cell carcinoma. *Comput. Biol. Med.* 100, 92–99.
- Xu, Z., Verma, A., Naveed, U., Bakhom, S., Khosravi, P., and Elemento, O. (2020). Using histopathology images to predict chromosomal instability in breast cancer: a deep learning approach. Preprint at medRxiv. <https://doi.org/10.1016/j.isci.2021.102394>.
- Luxton, J.J., McKenna, M.J., Lewis, A.M., Taylor, L.E., Jhavar, S.G., Swanson, G.P., and Bailey, S.M. (2021a). Telomere length and chromosomal instability for predicting individual radiosensitivity and risk via machine learning. *J. Personalized Med.* 11, 188.
- Li, X., Li, C., Rahaman, M.M., Sun, H., Li, X., Wu, J., Yao, Y., and Grzegorzec, M. (2022). A comprehensive review of computer-aided whole-slide image analysis: from datasets to feature extraction, segmentation, classification and detection approaches. *Artif. Intell. Rev.* 1–70.
- Chen, H., Li, C., Wang, G., Li, X., Mamunur Rahaman, M., Sun, H., Hu, W., Li, Y., Liu, W., Sun, C., et al. (2022). Gashis-transformer: a multi-scale visual transformer approach for gastric histopathological image detection. *Pattern Recogn.* 130, 108827.
- Coudray, N., Ocampo, P.S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., Moreira, A.L., Razavian, N., and Tsirigos, A. (2018). Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* 24, 1559–1567.
- Esteve, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., and Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444.
- Liu, L., Gao, Q., Jiang, J., Zhang, J., Song, X., Cui, J., Ye, Y., Wang, Z., Zhang, X., and Ren, X. (2020a). A survey on u-shaped networks in medical image segmentations. *Neurocomputing* 5, 244–258.
- Liu, L., Kurgan, L., Wu, F.-X., and Wang, J. (2020b). Attention convolutional neural network for accurate segmentation and quantification of lesions in ischemic stroke disease. *Med. Image Anal.* 65, 101791.
- Shmatko, A., Ghaffari Laleh, N., Gerstung, M., and Kather, J.N. (2022). Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. *Nat. Can. (Ott.)* 3, 1026–1038.
- Zhang, J., Li, C., Kosov, S., Grzegorzec, M., Shirahama, K., Jiang, T., Sun, C., Li, Z., and Li, H. (2021). Lcu-net: a novel low-cost u-net for environmental microorganism image segmentation. *Pattern Recogn.* 115, 107885.
- Chen, M., Zhang, B., Topatana, W., Cao, J., Zhu, H., Juengpanich, S., Mao, Q., Yu, H., and Cai, X. (2020b). Classification and mutation prediction based on histopathology h&e images in liver cancer using deep learning. *NPJ Precis. Oncol.* 4, 14–17.
- Fu, Y., Jung, A.W., Torne, R.V., Gonzalez, S., Vöhringer, H., Shmatko, A., Yates, L.R., Jimenez-Linan, M., Moore, L., and Gerstung, M. (2020). Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat. Can. (Ott.)* 1, 800–810.
- Xu, Z., Verma, A., Naveed, U., Bakhom, S.F., Khosravi, P., and Elemento, O. (2021). Deep learning predicts chromosomal instability from histopathology images. *iScience* 24, 102394.
- Bilal, M., Raza, S.E.A., Azam, A., Graham, S., Ilyas, M., Cree, I.A., Snead, D., Minhas, F., and Rajpoot, N.M. (2021). Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: a retrospective study. *Lancet. Digit. Health* 3, e763–e772.
- Ahmedt-Aristizabal, D., Armin, M.A., Denman, S., Fookes, C., and Petersson, L. (2021). A Survey on Graph-Based Deep Learning for Computational Histopathology (Computerized Medical Imaging and Graphics), p. 102027.
- Chen, R.J., Lu, M.Y., Shaban, M., Chen, C., Chen, T.Y., Williamson, D.F., and Mahmood, F. (2021). Whole slide images are 2d point clouds: context-aware survival prediction using patch-based graph convolutional networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer)*, pp. 339–349.
- Konda, R., Wu, H., and Wang, M.D. (2020). Graph convolutional neural networks to classify whole slide images. In *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, pp. 754–758.
- Zheng, Y., Jiang, Z., Xie, F., Shi, J., Zhang, H., Huai, J., Cao, M., and Yang, X. (2021). Diagnostic regions attention network (dr-net) for histopathology wsi recommendation and retrieval. *IEEE Trans. Med. Imag.* 40, 1090–1103.
- Zheng, Y., Jiang, Z., Zhang, H., Xie, F., Shi, J., and Xue, C. (2021). Histopathology wsi encoding based on gcn for scalable and efficient retrieval of diagnostically relevant regions. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2104.07878>.
- Gao, Z., Shi, J., and Wang, J. (2021). Gq-gcn: group quadratic graph convolutional network for classification of histopathological images. In *International Conference on Medical*

- Image Computing and Computer-Assisted Intervention (Springer), pp. 121–131.
31. Yang, W.-S., Feng, Y.-C., and Liang, T.-Y. (2021). An image quality predictor based on convolution neural networks. In 2021 International Symposium on Intelligent Signal Processing and Communication Systems (ISPAACS) (IEEE), pp. 1–2.
 32. Hendrycks, D., and Gimpel, K. (2016). Gaussian error linear units (gelus). Preprint at arXiv. <https://doi.org/10.48550/arXiv.1606.08415>.
 33. Nguyen, H.V., and Bai, L. (2010). Cosine similarity metric learning for face verification. In Asian conference on computer vision (Springer), pp. 709–720.
 34. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. (2017). Graph attention networks. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1710.10903>.
 35. Chen, D., Lin, Y., Li, W., Li, P., Zhou, J., and Sun, X. (2020a). Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. In Proceedings of the AAAI Conference on Artificial Intelligence, 34 Proceedings of the AAAI Conference on Artificial Intelligence, pp. 3438–3445.
 36. Scribano, C.M., Wan, J., Esbona, K., O’Sullivan, D.E., Lasek, A., McGregor, S.M., Fowler, A.M., Chappell, R.J., Bugni, T.S., Burkard, M.E., and Weaver, B.A. (2017). Integrative epigenetic and genetic pan-cancer somatic alteration portraits. *Epigenetics* 12, 561–574.
 37. Gutman, D.A., Khalilia, M., Lee, S., Nalisnik, M., Mullen, Z., Beezley, J., Chittajallu, D.R., Manthey, D., and Cooper, L.A.D. (2017). The digital slide archive: a software platform for management, integration, and analysis of histology for cancer research. *Cancer Res.* 77, e75–e78.
 38. Cheng, J., Gao, M., Liu, J., Yue, H., Kuang, H., Liu, J., and Wang, J. (2022). Multimodal disentangled variational autoencoder with game theoretic interpretability for glioma grading. *IEEE J. Biomed. Health Inform.* 26, 673–684.
 39. Ge, C., Gu, I.Y.-H., Jakola, A.S., and Yang, J. (2020). Deep semi-supervised learning for brain tumor classification. *BMC Med. Imag.* 20, 87.
 40. Li, Z., Wang, Y., Yu, J., Guo, Y., and Cao, W. (2017). Deep learning based radiomics (dlr) and its usage in noninvasive idh1 prediction for low grade glioma. *Sci. Rep.* 7, 5467–5511.
 41. Zhuge, Y., Ning, H., Mathen, P., Cheng, J.Y., Krauze, A.V., Camphausen, K., and Miller, R.W. (2020). Automated glioma grading on conventional mri images using deep convolutional neural networks. *Med. Phys.* 47, 3044–3053.
 42. Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 25.
 43. Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1409.1556>.
 44. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9.
 45. He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
 46. Deniz, E., Şengür, A., Kadiroğlu, Z., Guo, Y., Bajaj, V., and Budak, Ü. (2018). Transfer learning based histopathologic image classification for breast cancer detection. *Health Inf. Sci. Syst.* 6, 1–7.
 47. Granitto, P.M., Furlanello, C., Biasioli, F., and Gasperi, F. (2006). Recursive feature elimination with random forest for ptr-ms analysis of agroindustrial products. *Chemometr. Intell. Lab. Syst.* 83, 83–90.
 48. Yang, J., and Yang, J.-y. (2002). From image vector to matrix: a straightforward image projection technique: mpca vs. pca. *Pattern Recogn.* 35, 1997–1999.
 49. Xing, C., Ma, L., and Yang, X. (2016). Stacked denoising autoencoder based feature extraction and classification for hyperspectral images. *J. Sens.* 2016, 1–10.
 50. Ding, X., Xia, C., Zhang, X., Chu, X., Han, J., and Ding, G. (2021). Repmlp: Re-parameterizing convolutions into fully-connected layers for image recognition. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2105.01883>.
 51. Tolstikhin, I.O., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., Yung, J., Steiner, A., Keysers, D., Uszkoreit, J., et al. (2021). Mlp-mixer: an all-mlp architecture for vision. *Adv. Neural Inf. Process. Syst.* 34.
 52. Yu, T., Li, X., Cai, Y., Sun, M., and Li, P. (2021). S2-mlpv2: improved spatial-shift mlp architecture for vision. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2108.01072>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Python 3.7	Anaconda	https://www.anaconda.com
R	https://www.r-project.org	RRID: SCR 001905
HistomicsTK	Python library	Python library
Deposited data		
TCGA-BRCA	TCGA	https://portal.gdc.cancer.gov/
Other		
NVIDIA GPUs Nvidia	Corporation	https://www.nvidia.com/en-us/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Liangliang Liu (liangliu@henau.edu.cn).

Materials availability

The WSIs, clinical, and genomic data are available from TCGA-BRCA (<https://portal.gdc.cancer.gov/>). We strictly follow the TCGA data usage protocol.

Data and code availability

- The paper used existing datasets and code to train the computational models. The original datasets and codes are listed in the [key resources table](#).
- This study did not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Participants

The WSIs along with clinical and genomic data were downloaded from the TCGA-BRCA (<https://portal.gdc.cancer.gov/>). This dataset contains clinical, genetic, and pathological image data of patients, allowing data mining analysis by correlating cancer phenotype with genetic information. There are 1098 patients covering five carcinoma types, including 778 cases of ductal carcinoma.

(DC), 201 cases of lobular carcinoma, 16 cases of mucinous carcinoma, 8 cases of papillary carcinoma, and the remaining cases are mixed carcinoma. The formalin-fixed paraffin-embedded diagnostic H&E-stained histopathology WSIs from primary tumor sites are the subject of our study. After removing WSIs that lack magnification information, and/or WSIs with artifacts, including tissue folding, air bubbles, and out-of-focus regions, we get 896 cases with 896 WSIs, including 680 cases of ductal carcinoma, 200 cases of lobular carcinoma, and 16 cases of mucinous carcinoma.

METHOD DETAILS

Overview of MLP-GAT

As shown in [Figure 2](#), with the input WSI data, MLP-GAT architecture consists of seven parts. [Figure 2A](#) shows the tiles with 256×256 pixels. None of the tiles overlap and all of the tiles of each object come from the ROI in aWSI (Section 2.2 provides details). [Figure 2B](#) shows each tile is flattened into a tile vector with dimensions 256×256×3. There are n tiles in a WSI and each tile has three channels. [Figure 2C](#) shows

the MLP network used to extract dimensionally reduced tile vectors (Section 2.3 provides details). [Figure 2D](#) shows the extended layers for tumor classification, which is used to fine-tune the MLP method (Section 2.4 provides details). [Figure 2E](#) shows the determination of correlation values between any two tile vectors based on the CS method (Section 2.4 provides details). [Figure 2F](#) shows the representation of a constructed graph (Section 2.5 provides details). Finally, [Figure 2G](#) shows the proposed two-layer MLP-GAT and outputs the CIN status of the input WSI (Section 2.6 provides details).

Processing whole slide image

The sheer size and resolution of the WSI create severe challenges for image analysis algorithms. In this study, we draw on the experience of other scholars to slice the WSI. As shown in [Figure 2A](#), we divide the intermediate regions of each input WSI into 78 adjacent non-overlapping tiles. Each tile with dimensions $256 \times 256 \times 3$. To facilitate the input of the fully connected layer in the MLP module, we use the FlattenLayer method to compress each tile of the same WSI into a one-dimensional vector, which is shown in [Figure 2B](#).

Multi-layer perceptron

A MLP is a classical artificial neural network with forward structure, including input layer, output layer, and multiple hidden layers. The hidden layers of a MLP can learn the mapping relationship between the input layer and the output layer and realize transformation of feature dimensions. The layers of the MLP are fully connected.

Correlation analysis

We use the CS method to measure the correlation relationship between two vectors in the extracted vector matrix. The obtained correlation coefficients represent the edge weights between any two nodes in a graph. The CS method uses the cosine value of the angle between two vectors in vector space as a measure of the correlation coefficient between individuals.

Representation of constructed graph

The precise discrimination ability of the graph depends on the node values and edge weights. As shown in [Figure 2B](#), we construct the nodes and edges of a graph using the MLP and CS methods, respectively. In the constructed graph, the connection relationship between any two nodes is determined by the correlation coefficient of the corresponding tile feature vectors in a WSI. Here, each WSI is mapped onto a vector matrix of size 78×128 , where each WSI keeps 78 slices. Thus, there are 78 nodes in a graph. Based on Section 2.4, if the correlation coefficient between the vectors is greater than 0, the correlation coefficient is reserved as the edge weight between two nodes; otherwise, there is no connection between them.

MLP-GAT

The proposed MLP-GAT network is based on graph attention layers. MLP-GAT consists of three graph attention hidden layers. We perform the averaging operation on all node features in the last layer and make the classifying topics using the softmax activation function.

Experimental setup

We propose a correlation graph attention network (MLP-GAT) that can construct graphs for classifying multi-type CINs from the WSIs of breast cancer. We construct a WSIs dataset of breast cancer from the Cancer Genome Atlas Breast Invasive Carcinoma (TCGA-BRCA). There are three steps to train the entire pipeline. We first train the MLP and extended layers to produce tile vectors in the supervised manner. We then use the cosine similarity method to generate correlation coefficients. Finally, we construct graphs that are used to train and test the MLP-GAT. All the experiments are implemented in Pytorch on a papillary carcinoma with a NVIDIA GeForce Titan X Pascal CUDA GPU processor. To ensure fairness, these experiments were conducted in the training set and testing data with a ratio of 8:2.

Tile vector and correlation coefficient learning

We extract tile vectors for 896 cases. At this stage, 896 cases are divided training set and testing set (8:2), the training set is used to train and fine-tuning MLP and the extended layers, the testing set is used to verify the performance of this proposed method. We use the flattened tile vector of dimensions 196608×1 as the input for the MLP, we use the output vector of the second fully connected layer of the MLP as the extracted

feature of the input tile, with dimensions 128×1 . Finally, each WSI can be mapped into a vector matrix of size 78×128 . In this study, we use these vectors to construct nodes in the graph network. In addition, we use the CS method to measure the correlation coefficient between two tile vectors. We use these tile vectors and correlation coefficients to construct the graph on the WSI-level.

QUANTIFICATION AND STATISTICAL ANALYSIS

Significance results

We conduct MLP-GAT on the constructed graphs obtained from the 10X and 40X magnification WSIs, respectively. The 40X magnified tile-based MLP-GAT produced a better classification than that 10X magnified tile-based MLP-GAT, mainly because the 40X magnified tiles offer much more detail of the pathological tissues than the 10X magnified tiles. The results of MLP-GAT were compared with that of the other advanced methods. To demonstrate the effectiveness of feature vectorization and similarity coefficient methods on MLP-GAT, we undertake several experiments on the similarity coefficient between feature vectors. In addition, we take the WSI of a ductal carcinoma sample as an example and selected 64 tiles for subtype prediction probability analysis to verify the heterogeneity among tiles in the same WSI. The statistical results are reported alongside their descriptions in the main text of the paper.