Integrative Analysis of Multi-view Histopathological Image Features for the Diagnosis of Lung Cancer

Zongxiang Pei, Yingli Zuo, Liang Sun, Meiling Wang, Daoqiang Zhang(✉), and Wei Shao(✉)

MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China
{dqzhang,shaowei20022005}@nuaa.edu.cn

Abstract. Lung cancer is one of the most widely spread cancers in the world. So far, the histopathological image remains the "gold standard" in diagnosing lung cancers, and multiple types of pathological images features have been associated with lung cancer diagnosis and progression. However, most of the existing studies only utilized single type of image features, which did not take advantages of multiple types of image features. In this paper, we propose a Block based Multi-View Graph Convolutional Network (*i.e.*, BMVGCN), which integrates multiple types of image features from histopathological images for lung cancer diagnosis. Specifically, our method utilizes the block-based bilinear combination model to fuse different types of features. By considering the correlation among different samples, we also introduce the Graph Convolutional Network to exploit the correlations among samples that could lead to better diagnosis performance. To evaluate the effectiveness of the proposed method, we conduct the experiments for the classification of the cancer tissue and non-cancer tissue in both Lung Adenocarcinoma (*i.e.*, LUAD) and Lung Squamous Cell Carcinoma (*i.e.*,LUSC), and the discrimination between LUAD and LUSC. The results show that our method can achieve superior classification performance than the comparing methods.

Keywords: Lung cancer diagnosis · Histopathological image · Graph neural network · Multi-view fusion

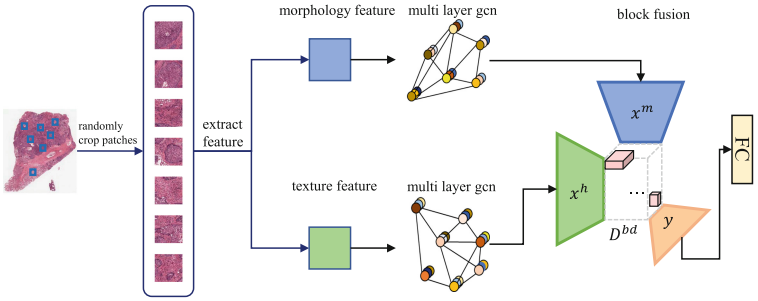
1 Introduction

1.1 Related Work

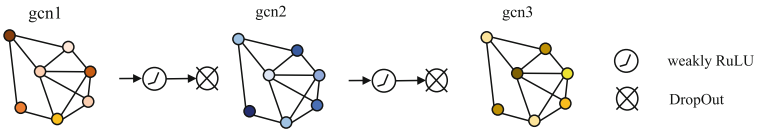
Nowadays, cancer can be diagnosed through multiple imaging biomarkers, including Computed Tomography (CT), Positron Emission Tomography/Computed

This work was supported by National Natural Science Foundation of China (Nos. 62136004, 61902183, 62106104), the National Key R&D Program of China (Grant Nos.: 2018YFC2001600, 2018YFC2001602 and 2018YFA0701703), and the Project funded by China Postdoctoral Science Foundation (No. 2022T150320).

L. Fang et al. (Eds.): CICAI 2022, LNAI 13605, pp. 577–587, 2022.
https://doi.org/10.1007/978-3-031-20500-2_47



(a) The framework of BMVGCN.



(b) Details of multi layer gcn.

Fig. 1. The overall framework of the proposed method.

Tomography (PET/CT), MRI and histopathological images. Among all these imaging biomarkers, histopathological images are generally considered to be the gold standard for cancer diagnosis and prognosis since it can provide morphological attributes of cells that are highly related to the degree of the aggressiveness of cancers [1]. With the help of ever-increasing computing resources, many computational histopathological systems have been proposed to extract different types of histopathological image features to help diagnose human cancers [2–7]. For example, Gurcan et al. [3] utilized Gray-level co-occurrence matrix (GLCM) and Gray-level run length matrix (GLRLM) to extract texture features (*i.e.*, haralick features) from the histopathological images, which are proved to be sensitive to the diagnosis of brain cancer, Sparks et al. [6] presented a set of explicit shape descriptors (ESDs) to obtain morphology features for classifying gland Gleason grade in prostate cancers. Other studies include Shukla et al. [7] have extracted morphological features for accurate and reliable detection of breast cancers.

Recently, besides applying single type of image features for cancer diagnosis, several studies combined multiple types of image features for diagnosing human cancers. For instance, Cheng et al [8] proposed an 150-dimensional feature including both morphological and color information in histopathological image for diagnosing kidney cancer. Yu et al. [9] demonstrated that the combination of morphology and texture features can better predict the prognosis of lung cancer patients than using single type of features. Although the above methods indicated that the combination analysis of multiple types of image features can uncover the hidden difference between normal and cancer or different cancer subtypes that cannot be found using single type of features. Most of the studies

directly combined different types of image features for the diagnosis task, which neglected the weight information for each feature type in the fusion process. As a matter of fact, different types of features may carry different task-relevant information, and fusing them by naive concatenation may limit the model's ability to dynamically determine the relevance of each type of features for the cancer diagnosis task. In addition, most of the integrative models only considered the correlation within the multiple types of features, and thus neglected to take the association among different samples into consideration that will deteriorate the cancer diagnosis performance.

Based on the above considerations, in this paper, we integrate multiple types of image features (*i.e.*, morphology and texture features) from histopathological images and propose a Block based Multi-View Graph Convolutional Network (*i.e.*, BMVGCN) for lung cancer diagnosis. Specifically, our model utilizes the block-based bilinear combiantion model [10] to fuse different types of features, which aims at automatically learning the weight for the combination of different types of image features. In addition, to better exploit the association among different samples to help lung cancer diagnosis, we introduce the Graph Convolutional Network to exploit their correlations through similarity networks. To evaluate the effectiveness of the proposed method, we conduct the experiments for the classification of the cancer tissue and non-cancer tissue in both Lung Adenocarcinoma (*i.e.*, LUAD) and Lung Squamous Cell Carcinoma (LUSC), and the discrimination between LUAD and LUSC. The results show that our method can achieve superior classification performance than the comparing methods.

2 Methods

2.1 Datasets

All the histopathological images were collected from Nanjing Medical University. Patients who received any treatment or neoadjuvant therapy before surgery/biopsy were excluded. Samples (tumor specimens, adjacent normal tissues, and peripheral blood) were obtained during surgical resection. All tissue samples were snap-frozen. HE-stained sections from each sample were subjected to an independent pathology review to confirm that the tumor specimen was histopathologically consistent with NSCLC (>70% tumor cells). For Lung Adenocarcinoma (LUAD) cohort, it contains 73 cancer and 163 normal samples. As to LUSC cohort (Lung Squamous Cell Carcinoma), 53 cancer samples and 61 normal samples are involved. We show the demographics information of these two cohorts in Table 1.

Table 1. The demographic information for different lung cancer cohorts

	Male/Female	Age	Tumor/Nontumor
LUAD	115/121	58.65 ± 10.67	73/163
LUSC	100/14	61.25 ± 8.36	53/61

2.2 Overview of Our Method

We summarize our framework in Fig. 1, which consists of the following four steps. Firstly, we extract different types of image features (*i.e.*, morphology and texture features) from pathological images. Secondly, for each type of features, we apply graph convolutional network (*i.e.*, GCN) to learn their high-level representation through sample similarity network. Thirdly, we apply the block-based bilinear combination model to fuse different types of features. Finally, we feed the fused feature into fully connected layers for the final classification task.

2.3 Feature Extraction

Before feature extraction procedure, the patches are cropped from wsi images. For each sample, we randomly crop 10 to 20 patches. Those patches with background area (brightness more than 78%) less than 30% are retained.

For morphology features, we firstly apply an unsupervised method introduced in [11] to segment the nuclei from the raw histopathological image. The utilized cell segmentation algorithm is comprised of three steps. Firstly, the color deconvolution operation is adopted to derive the gray-scale image in hematoxylin channel of the input H/E stained histopathological images. Then, the resulting grayscale image is processed with opening by reconstruction to connect close background regions to each other. Finally, a multi-level thresholding segmentation method, whose threshold can be automatically adjusted according to each input image is presented for the segmentation of cells. Then, for each segmented nucleus with their area ranging from 10 to 200, we extract five cell-level features characterizing the nuclei area (denoted as area), the major and minor axis length of cell nucleus (major and minor), the eccentricity of the nucleus (eccentricity), the ratio of major axis length to minor axis (ratio). After that, for each cell-level feature, we summarize all cell-level features into sample-level features by using a 10-bin histogram and five statistical measurements (*i.e.*, mean, standard deviation, skewness, kurtosis, and entropy). Thus, 75-dimensional morphology features can be derived for each sample. We use the same naming rule for both cell-level and sample-level features. For instance, the feature major_bin_1 represents the percentage of cells with small major axis length while major_bin_10 referred to the percentage of nuclei with long major axis. As to texture features, we extract haralick features from the histopathological images. Then, we extract 13-dimensional haralick features (i.e., Hara_1, Hara_2, ..., Hara_13) for each valid patch. Like the aggregation method for morphology features, we summarize all patch-level haralick features into 195-dimensional features for each sample. For instance, Hara_1_bin_1, Hara_1_bin_2,..., Hara_1_bin_10 represent the ten histogram features for Hara_1, and Hara_1_skewness indicates the skewness feature for Hara_1.

2.4 High-Level Feature Learning by Graph Convolutional Network

With the consideration of the correlation among different subjects can promote the classification performance, we utilize Graph Convolutional Network (*i.e.*, GCN) to

extract high-level representation from each type of extracted features. The GCN model is comprised of two main parts, the first part is the node feature matrix and the second part is an adjacency matrix which can be used to describe the structure of graph.

Table 2. Results for LUAD *v.s.* nonLUAD and LUSC *v.s.* nonLUSC.

Task	Measurements	ACC	F1-Score	AUC	Recall	Precision
LUAD v.s. nonLUAD	RF	0.758	0.713	0.911	0.274	0.833
	SVML1	0.860	0.740	0.921	0.644	0.870
	SVML2	0.805	0.646	0.851	0.575	0.737
	SALMON	0.869	0.783	0.643	0.767	0.800
	GCN_BlockTucker	0.907	0.905	0.924	0.907	0.906
	GCN_Mutan	0.919	0.919	0.921	0.919	0.919
	GCN_MFH	0.919	0.918	0.909	0.919	0.919
	BMVGCN	**0.928**	**0.927**	**0.953**	**0.928**	**0.927**
LUSC v.s. nonLUSC	RF	0.877	0.865	0.934	0.849	0.882
	SVML1	0.596	0.681	0.716	0.925	0.538
	SVML2	0.868	0.860	0.958	0.868	0.852
	SALMON	0.754	0.781	0.720	0.943	0.667
	GCN_BlockTucker	0.895	0.895	0.948	0.895	0.895
	GCN_Mutan	0.939	0.938	0.879	0.939	0.940
	GCN_MFH	0.930	0.930	0.895	0.930	0.930
	BMVGCN	**0.947**	**0.947**	**0.981**	**0.947**	**0.947**
LUAD v.s. LUSC	RF	0.571	0.727	0.849	**0.986**	0.576
	SVML1	0.563	0.715	0.560	0.945	0.575
	SVML2	0.722	0.780	0.783	0.849	0.721
	SALMON	0.746	0.775	0.715	0.753	0.797
	GCN_BlockTucker	0.881	0.882	0.921	0.881	0.891
	GCN_Mutan	0.873	0.874	0.857	0.873	0.876
	GCN_MFH	0.881	0.881	0.849	0.881	0.883
	BMVGCN	**0.897**	**0.897**	**0.934**	0.897	**0.904**

Node Feature Matrix. The GCN model is comprised of two components, the first component is the node feature matrix $\boldsymbol{X}^t \in \mathbb{R}^{n \times d_t}, t \in \{m, h\}$. Specifically, let $\boldsymbol{X}^m \in \mathbb{R}^{n \times d_m}$ and $\boldsymbol{X}^h \in \mathbb{R}^{n \times d_h}$ be the extracted morphology and haralick texture features, respectively. Here, n represents the sample size, and d_m and d_h correspond to the dimensionality of the morphology and haralick texture features, respectively. By viewing each sample as a node in sample similarity network, the goal of applying GCN is to learn the function of each type of features on a graph to obtain high-level features that can capture the correlation among different samples.

Adjacent Matrix. The second component in GCN model is used to describe the graph structure, which can be represented in the form of an adjacency matrix $\boldsymbol{A}^t \in \mathbb{R}^{n\times d}, t \in \{m, h\}$. Our proposed method contains two types of features (*i.e.*, texture features and morphological features), and for each type of features, we use a graph to depict the correlation among different samples. We denote them $G_m = \{\boldsymbol{V^m}, \boldsymbol{A^m}\}$ and $G_h = \{\boldsymbol{V^h}, \boldsymbol{A^m}\}$ respectively. In the graph $G_t, t \in \{m, h\}$, each node represents a sample, the initialized adjacency matrix $\boldsymbol{A}^t$ for feature type t in GCN is constructed by calculating the cosine similarity between pairs of nodes. To control the number of edges in the adjacency matrix $\boldsymbol{A}^t$, we introduce a variable ε as a threshold, and edges with larger cosine similarity than ε are retained. Then, the adjacency between node i and node j in graph $\boldsymbol{A}^t$ can be calculated as:

$$\boldsymbol{A}^t_{ij} = \begin{cases} s(\boldsymbol{x}^t_i, \boldsymbol{x}^t_j), & \text{if } i \neq j \text{ and } s(\boldsymbol{x}^t_i, \boldsymbol{x}^t_j) \geq \varepsilon \\ 0, & otherwise \end{cases} \tag{1}$$

where $\boldsymbol{x}^t_i$ and $\boldsymbol{x}^t_j$ are the representations of node i and node j for feature type t, respectively. $s(\boldsymbol{x}_i, \boldsymbol{x}_j) = \frac{\boldsymbol{x}_i \cdot \boldsymbol{x}_j}{\|x_i\|_2 \|x_j\|_2}$ is the cosine similarity between node i and j. The threshold ε can be determined by a parameter k, which represents the average number of edges per node that are retained except self loops:

$$k = \sum_{i,j,i\neq j} \boldsymbol{I}(s(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq \varepsilon)/n \tag{2}$$

where $\boldsymbol{I}(\cdot)$ is the indicator function and n is the number of nodes.

Construction of Graph Convolutional Network. A GCN is built by multiple convolutional layers, and each layer of specific feature type is defined as:

$$\begin{aligned} \boldsymbol{H}^t_{(l+1)} &= f(\boldsymbol{H}^t_{(l)}, \boldsymbol{A}^t_{(l)}) \\ x'x &= \sigma(\boldsymbol{A}^t_{(l)} \boldsymbol{H}^t_{(l)} \boldsymbol{W}^t_{(l)}), \end{aligned} \tag{3}$$

where $\boldsymbol{H}^t_{(l)}$ is the input of the l-th layer for feature type t, $\boldsymbol{W}^t_{(l)}$ and $\boldsymbol{A}^t_{(l)}$ refer to its corresponding weight and adjacent matrix. $\sigma(\cdot)$ denotes the non-linear activation function. In the training procedure of GCN, we follow the method introduced in [12] and modify the adjacency matrix $\boldsymbol{A}^t$ as:

$$\widetilde{\boldsymbol{A}}^t_{(l)} = (\hat{\boldsymbol{D}}^t_{(l)})^{-\frac{1}{2}} (\boldsymbol{A}^t_{(l)} + \boldsymbol{I}^t_{(l)}) (\hat{\boldsymbol{D}}^t_{(l)})^{-\frac{1}{2}}, \tag{4}$$

where $\hat{\boldsymbol{D}}^t_{(l)}$ is the diagonal node degree matrix of $\hat{\boldsymbol{A}}^t_{(l)}$ for feature type t and $\boldsymbol{I}^t_{(l)}$ represents the identity matrix. Then, we denote the output of the GCN with L layers as:

$$\boldsymbol{Y}^t_{(L)} = GCN_L(\boldsymbol{X}^t, \widetilde{\boldsymbol{A}}^t) \tag{5}$$

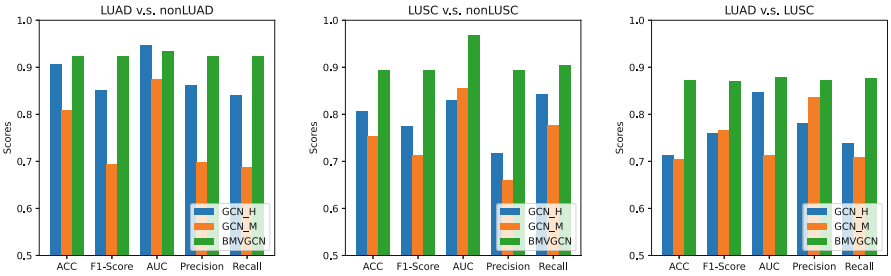


Fig. 2. The classification results by combining different types of features and using single type of features

2.5 The Block-Based Bilinear Combination Model

After applying GCN for feature learning, we derive high-level representation for both morphology and haralick features. Here, we denote the morphology and texture feature for the u-th sample as $x^{(m,u)}$ and $x^{(h,u)}$, respectively. Since different types of features may provide complementary information for the following diagnosis task, we apply block based bilinear combination model to fuse different types of features [13–17]. Specifically, let $x^{(m,u)} \in \mathbb{R}^I$ and $x^{(h,u)} \in \mathbb{R}^J$ be the input high-level morphology and haralick features of the block based bilinear combination model. The block based bilinear combination model can combine $x^{(m,u)}$ and $x^{(h,u)}$ into a K-dimensional output vector with tensor products:

$$\boldsymbol{y}^u = \tau \times_1 x^{(m,u)} \times_2 x^{(h,u)} \tag{6}$$

where $\boldsymbol{y}^u \in \mathbb{R}^K$, and $\times_1$ and $\times_2$ represents tensor products respectively. Each component of y^u $(i.e., y_k^u)$ is a quadratic form of the inputs: $\forall K \in [1, K]$,

$$y_k^u = \sum_{i=1}^{I} \sum_{j=1}^{J} \tau_{ijk} \cdot x_i^m \cdot x_j^h \tag{7}$$

where I and J indicate the diemnsionality of $x^{(m,u)}$ and $x^{(h,u)}$ respectively. A bilinear model is completely defined by its associated tensor $\tau \in \mathbb{R}^{I \times J \times K}$. In order to reduce the number of parameters and constrain the model's complexity, we introduce Block model [10], which applied the block-term decomposition method to obtain τ. The decomposition of τ is defined as:

$$\tau := \sum_{r=1}^{R} \boldsymbol{D} \times_1 \boldsymbol{A}_r \times_2 \boldsymbol{B}_r \times_3 \boldsymbol{C}_r \tag{8}$$

where $\boldsymbol{A}_r \in \mathbb{R}^{I \times L}$, $\boldsymbol{B}_r \in \mathbb{R}^{J \times M}$ and $\boldsymbol{C}_r \times \mathbb{R}^{K \times N}$. $\times_1$, $\times_2$ and $\times_3$ represent the tensor products. The block-term decomposition of τ can be formulated as:

$$\tau = \boldsymbol{D}^{bd} \times_1 \boldsymbol{A} \times_2 \boldsymbol{B} \times_3 \boldsymbol{C} \tag{9}$$

where $\boldsymbol{A} = [\boldsymbol{A}_1, ..., \boldsymbol{A}_R]$ (same for $\boldsymbol{B}$ and $\boldsymbol{C}$), and $\boldsymbol{D}_{bd} \in \mathbb{R}^{LR \times MR \times NR}$ is the block-superdiagonal tensor of $\{\boldsymbol{D}_r\}_{1 \leq r \leq R}$. Let $\hat{x}^{(m,u)} = \boldsymbol{A}x^{(m,u)} \in \mathbb{R}^{LR}$ and $\hat{x}^{(h,u)} = \boldsymbol{B}x^{(h,u)} \in \mathbb{R}^{MR}$,we can fuse different types of features by the block-superdiagonal tensor $\boldsymbol{D}^{bd}$. Each block in $\boldsymbol{D}^{bd}$ merges chunks from $\hat{x}^{(m,u)}$ and $\hat{x}^{(h,u)}$ to generate z_r^u with size N:

$$z_r^u = \boldsymbol{D}_r \times_1 \hat{x}^{(m,u)}_{rL:(r+1)L} \times_2 \hat{x}^{(h,u)}_{rM:(r+1)M} \tag{10}$$

where $\hat{x}^{(q,u)}_{i:j} q \in \{m, h\}$ is a vector of dimension $j - i$. After concatenating all the z_r to generate $z^u \in \mathbb{R}^{NR}$. The output of the block-based bilinear combination model can be calculated by $y^u = Cz^u \in \mathbb{R}^K$.

After applying the block based bilinear combination model to integrate different types of features, a two-layer fully-connected neural network followed by the softmax function is applied to predict the label of each sample.

3 Results

3.1 Experimental Results

Our proposed method contains three graph convolutional layers. As can be seen from Fig. 1b, weakly relu layer and dropout layer are added after the first and the second graph convolutional layers. The relu ratio and dropout ratio are set as 0.25 and 0.5 respectively. The number of adjacent edges in GCN is set as 2. The dimensionality of the hidden layer of the fully connected layer is set as 100. Total epochs for training different classification models are set as 100. We use the leave-one-out strategy to evaluate the performance of different methods by the measurements of recall, precision, auc, and the f1-score. For all the samples in the training dataset, we randomly split 20% of them as validation set to tune the model hyperparameters. The experiments are conducted on a computer with 32-GB memory, Intel I9-10900X 3.7 GHz CPU, and NVIDIA GeForce RTX 3090 GPU. Moreover, the proposed method and all neural-network-based baseline models are implemented based on PyTorch 1.8.1.

3.2 Intergrating Two Types of Features Performs Better Than Only Using One Type

We first investigate the effect of using single type of features (*i.e.*, GCN_M and GCN_H) and integrating different types of features together (*i.e.*, BMVGCN) for the diagnosis of lung cancer. Here, GCN_M and GCN_H refer to the methods that only applying morphological and haralick features followed by the GCN for the classification task, respectively. We test the performance of different methods on the following three different tasks.

– LUAD *v.s.* nonLUAD: Classify LUAD samples (denoted as LUAD) and their corresponding normal samples (denoted as nonLUAD).

- LUSC *v.s.* nonLUSC: Classify LUSC samples (denoted as LUSC) and their corresponding normal samples (denoted as nonLUSC).
- LUAD *v.s.* LUSC: Classify LUAD tumor samples (LUAD) and LUSC tumor samples (LUSC).

As can be seen from Fig. 2, the combination of different types of features (*BMVGCN*) can better diagnose lung cancer than its competitors that only applying single type of features on all the classification tasks. These results strongly validate the effectiveness of integrating different types of features in distinguishing tumor and non-tumor tissues and the classification of different lung cancer subtypes. In addition, we note that GCN_H is generally superior to GCN_M across all tasks, which demonstrates that the texture features play a more important role in the classification tasks.

3.3 Comparison of *BMVGCN* and Other Methods for Lung Cancer Classification

In Sect. 3.2, We have shown the effectiveness of combining different types of features for lung cancer diagnosis. To further verify the superiority of our proposed method, we compare BMVGCN with the following 7 multi-view learning algorithms for the lung cancer diagnosis.

- RF: Concatenate haralick and morphology features at first, and then feeds the concatenated features into random forest classifier to obtain the predicted label,
- SVML1: Feed the concatenated features into support vector machine with l1 loss function,
- SVML2: Feed the concatenated features into support vector machine with l2 loss function,
- SALMON: A deep learning based multi-view learning algorithm proposed in [18],
- GCN_BlockTucker: This module correponds to Block without the low-rank constraint on third-mode slices of D_c tensors [10],
- GCN_Mutan: A GCN based multi-view learning algorithm proposed in [19],
- GCN_MFH : A GCN based multi-view learning algorithm proposed in [20].

Similar to Sect. 3.2, we conduct experiments on LUAD *v.s.* nonLUAD, LUSC *v.s.* nonLUSC and LUAD *v.s.* LUSC these three tasks. Table 2 shows the classification results including accuracy (ACC), F1-Score (f1-score), AUC (Area under the ROC curve), Recall and Precision. From Table 2, we can derive the following three observations: 1) The performance of deep learning based methods (*i.e.*, SALMON, GCN_BlockTucker, GCN_Mutan, GCN_MFH, BMVGCN) are superior to traditional machine learning algorithms (*i.e.*,RF, SVML1, SVML2) across all tasks, which reveals the advantages of applying deep learning algorithms for the classification of lung cancers. 2) The GCN based algorithms can better classify lung cancer patients since they consider the correlations among different patients for better representation of the input features. 3) The proposed

Table 3. Ablation study results of LUAD *v.s.* nonLUAD and LUSC *v.s.* nonLUSC and LUAD *v.s.* LUSC

	Method	ACC	F1-Score	AUC	Recall	Precision
LUAD v.s. nonLUAD	(A) AFF + FC	0.831	0.835	0.888	0.831	0.852
	(B) GCN + FC	0.860	0.852	0.911	0.860	0.865
	(C) BMVGCN	0.928	0.927	0.953	0.928	0.927
LUSC v.s. nonLUSC	(A) AFF + FC	0.833	0.830	0.819	0.833	0.849
	(B) GCN + FC	0.886	0.886	0.957	0.886	0.886
	(C) BMVGCN	0.947	0.947	0.981	0.947	0.947
LUAD v.s. LUSC	(A) AFF + FC	0.762	0.763	0.778	0.762	0.771
	(B) GCN + FC	0.786	0.785	0.880	0.786	0.785
	(C) BMVGCN	0.897	0.897	0.934	0.897	0.904

fusion algorithm can consistently obtain superior classification results than the comparing methods, which reveals the fact that the block model embedded in our method is effective since using block model can not only automatically fuse different types of features, but also reduce the complexity of the whole model.

3.4 Ablation Study

To evaluate the effectiveness of different components in the proposed method, we conduct ablation studies. Here, we conduct experiments on the following three configurations: (A) AFF + FC: Directly integrates haralick texture features and morphology features by block based bilinear combination model without applying GCN to learn high-level representations. (B) GCN + FC: Applying fully connected layers to integrate high-level haralick texture and morphology features without block based bilinear combination model after GCN. (C) *BMVGCN*: our proposed method. The results of all these three methods are shown in Table 3.

As can be seen from Table 3, our proposed method consistently achieves better classification performance than its competitors. These results show the necessity of applying GCN and block based bilinear combination model for the diagnosis of lung cancer from histopathological images. In addition, comparing (B) to (A) in all three tasks, it is worth noting that GCN + FC outperforms AFF + FC on all measurements. This demonstrates that GCN plays a more important role for the classification of lung cancer in comparison with block based bilinear combination model.

References

1. Rubin, R., Strayer, D.S., Rubin, E., et al.: Rubin's Pathology: Clinicopathologic Foundations of Medicine. Lippincott Williams & Wilkins (2008)
2. Belsare, A., Mushrif, M., Pangarkar, M., Meshram, N.: Classification of breast cancer histopathology images using texture feature analysis. In: TENCON 2015–2015 IEEE Region 10 Conference. IEEE, 2015, pp. 1–5 (2015)

3. Durgamahanthi, Vaishali, Anita Christaline, J.., Shirly Edward, A..: GLCM and GLRLM based texture analysis: application to brain cancer diagnosis using histopathology images. In: Dash, Subhransu Sekhar, Das, Swagatam, Panigrahi, Bijaya Ketan (eds.) Intelligent Computing and Applications. AISC, vol. 1172, pp. 691–706. Springer, Singapore (2021). https://doi.org/10.1007/978-981-15-5566-4_61
4. de Matos, J., de Souza Britto, A., de Oliveira, L.E., Koerich, A.L.: Texture CNN for histopathological image classification. In: IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS). IEEE, pp. 580–583 (2019)
5. Roncalli, M., Park, Y.N., Di Tommaso, L.: Histopathological classification of hepatocellular carcinoma. Dig. Liver Dis. **42**, S228–S234 (2010)
6. Sparks, R., Madabhushi, A.: Explicit shape descriptors: novel morphologic features for histopathology classification. Med. Image Anal. **17**(8), 997–1009 (2013)
7. Shukla, K., Tiwari, A., Sharma, S., et al.: Classification of histopathological images of breast cancerous and non cancerous cells based on morphological features. Biomed. Pharmacol. J. **10**(1), 353–366 (2017)
8. Cheng, J., et al.: Integrative analysis of histopathological images and genomic data predicts clear cell renal cell carcinoma prognosis. Can. Res. **77**(21), e91–e100 (2017)
9. Yu, K.M., et al.: Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. Nat. Commun. **7**(1), 1–10 (2016)
10. Ben-Younes, H., Cadene, R., Thome, N., Cord, M.: Block: Bilinear superdiagonal fusion for visual question answering and visual relationship detection. Proc. AAAI Conf. Artif. Intell. **33**(01), 8102–8109 (2019)
11. Phoulady, H.A., Goldgof, D.B., Hall, L.O., Mouton, P.R.: Nucleus segmentation in histology images with hierarchical multilevel thresholding. In: Medical Imaging 2016: Digital Pathology, vol. 9791. International Society for Optics and Photonics, p. 979111 (2016)
12. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
13. Gao, Y., Beijbom, O., Zhang, N., Darrell, T.: Compact bilinear pooling. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 317–326 (2016)
14. Lin, T.-Y., Maji, S.: Improved bilinear pooling with CNNs. arXiv preprint arXiv:1707.06772 (2017)
15. Fukui, A., Park, D.H., Yang, D., Rohrbach, A., Darrell, T., Rohrbach, M.: Multimodal compact bilinear pooling for visual question answering and visual grounding. arXiv preprint arXiv:1606.01847 (2016)
16. Yu, Z., Yu, J., Fan, J., Tao, D.:Multi-modal factorized bilinear pooling with co-attention learning for visual question answering. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1821–1830 (2017)
17. Zhang, Y., Tang, S., Muandet, K., Jarvers, C., Neumann, H.: Local temporal bilinear pooling for fine-grained action parsing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12 005–12 015 (2019)
18. Huang, Z., et al.: Salmon: survival analysis learning with multi-omics neural networks on breast cancer. Front. Genet. **10**, 166 (2019)
19. Ben-Younes, H., Cadene, R., Cord, M., Thome, N.: Mutan: Multimodal tucker fusion for visual question answering. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2612–2620 (2017)
20. Shah, N., Singhal, N., Singh, C., Khandelwal, Y.:Model agnostic information biasing for VQA. In: 8th ACM IKDD CODS and 26th COMAD, 2021, pp. 419–419 (2021)