Predicting the Survival of Cancer Patients With Multimodal Graph Neural Network

Jianliang Gao, Tengfei Lyu, Fan Xiong, Jianxin Wang, Weimao Ke, and Zhao Li

Abstract—In recent years, cancer patients survival prediction holds important significance for worldwide health problems, and has gained many researchers attention in medical information communities. Cancer patients survival prediction can be seen the classification work which is a meaningful and challenging task. Nevertheless, research in this field is still limited. In this work, we design a novel Multimodal Graph Neural Network (MGNN) framework for predicting cancer survival, which explores the features of real-world multimodal data such as gene expression, copy number alteration and clinical data in a unified framework. Specifically, we first construct the bipartite graphs between patients and multimodal data to explore the inherent relation. Subsequently, the embedding of each patient on different bipartite graphs is obtained with graph neural network. Finally, a multimodal fusion neural layer is proposed to fuse the medical features from different modality data. Comprehensive experiments have been conducted on real-world datasets, which demonstrate the superiority of our modal with significant improvements against state-of-the-arts. Furthermore, the proposed MGNN is validated to be more robust on other four cancer datasets.

Index Terms—Medical information retrieval, cancer survival prediction, graph neural networks, multimodal

1 Introduction

Nowadays, graph neural networks (GNN) [1] has great practical performance of learning representation for graph-based tasks, and is being generated across multiple high-impact application domains, including but not limited to bioinformatics [2], social network analysis [3] and recommendation [4]. In particular, biological networks[5], [6], [7], [8] analysis and discovery based on deep learning plays an important role in public health [9], and has gained many researchers attention in many specific applications such as drug-drug interaction prediction [10], disease type prediction[11] and epidemiological prediction [12].

Inspired by the success of graph neural networks in graph-based data, it is natural to think that GNN have the potential to improve the effectiveness of predicting the survival of cancer patients in bioinformatics area. Although predicting cancer prognosis has remarkable meaning after cancer disease has become one of the most important worldwide health problems. Additionally, cancer patients survival expectancy prediction is a fundamental task among predicting cancer prognosis [13]. Predicting cancer survival can be formulated as a censored survival analysis classification problem, predicting cancer patients both if and when an event (i.e., death) occurs within a given time period [14].

- *Jianliang Gao, Tengfei Lyu, Fan Xiong, and Jianxin Wang are with the School of Computer Science and Engineering, Central South University, Changsha, Hunan 410083, China. E-mail: {gaojianliang, tengfeilyu, xiongfan, jxwang}@csu.edu.cn.*
- *Weimao Ke is with the College of Computing & Informatics, Drexel University, Philadelphia, PA 19104 USA. E-mail: wk@drexel.edu.*
- *Zhao Li is with the Alibaba Group, Hangzhou, Zhejiang 311121, China. E-mail: lizhao.lz@alibaba-inc.com.*

Manuscript received 31 Aug. 2020; revised 11 Mar. 2021; accepted 12 May 2021. Date of publication 25 May 2021; date of current version 1 Apr. 2022. (Corresponding author: Jianxin Wang.)
Digital Object Identifier no. 10.1109/TCBB.2021.3083566

In medical cancer disease classification area, the long term survival rate and the short term survival rate are two common used indexes for cancer prognosis. For example, among breast cancer patients, 5-year is usually used index. Patients with a survival time of more than five years are marked as long term survival, and less than five years are marked as short term survival. Only slightly more than half of the newly diagnosed oral cancer patients will survive over the long term survival [15].

Cancer is one of the major worldwide health problems [16]. Diagnosis and detection of cancer disease in early stage is a strong sense of purpose for cancer cure. Over the past years, a continuous evolution of cancer research has been performed. Among the diverse methods and techniques developed for cancer prediction, one of the most important research hotspot is to utilize the gene expression level in this field. Data analysis on gene expression level has facilitated cancer diagnosis and treatment to a great extent. Cancer prognosis prediction is one of the most critical and urgent tasks for physicians [17]. For example, breast cancer has become the malignant tumor with the highest survival rate after the thyroid tumor. Breast cancer at early stage can be completely cured, and patients with middle and advanced stage can also achieve long-term survival with tumor through targeted therapy, endocrine therapy and other systemic therapies. It is well known that genes and diseases are closely related[18], [19], and it is absolutely meaningful to fully explore the hidden information in genes for cancer prediction. Also, cancer is particularly complex and it is extremely difficult to predict and treat [20]. Therefore, cancer prognosis prediction plays an important role in clinical works.

Accurate prediction of the cancer survival could help clinicians make effective decisions and establish appropriate therapy programs [21]. Meanwhile, it can spare a significant number of patients from receiving unnecessary treatment and its related expensive medical costs [22]. Many researchers have

attempted to solve the task of predicting cancer prognosis. Support vector machine-based recursive feature elimination approach is proposed for prognosis prediction by relying on gene expression data [23]. Sun *et al.* propose a multimodal deep neural network by integrating multi-dimensional data for predicting breast cancer prognosis [22]. Recently, deep learning has been an emerging methodology and provided a possible approach to improve the accuracy of predicting cancer survival [24]. Although the above deep learning-based methods have achieved acceptable performance on cancer survival prediction, there still exists significant room for improvement, especially in the following two aspects: (1) How to utilize the structure information between patients and multimodal medical data. Most of previous works ignore the inherent relations between patients and medical data. (2) How to utilize the features of medical data from different modalities (e.g., gene expression and clinical). Most of prior works directly combine different types of data into deep learning model without utilizing the rich information of multimodal data.

To tackle these challenges, in this work, we further propose a novel Multimodal Graph Neural Network (MGNN) for cancer patients survival prediction. We first construct a bipartite graph utilizing gene expression profile or the DNA copy number alteration (CNA) profile, which explores the inherent relation among them. Based on the bipartite graphs, we propose a new graph neural network to obtain the embedding representation of each patient. Finally, the output is the classification of short term survival or long term survival based on a given threshold (breast cancer usually adopt 5-year survival in the community). In summary, this paper has the following contributions:

- We highlight the critical importance of explicitly exploiting the multimodal data, and the inherent relation between cancer patients and multimodal data.
- We design a novel Multimodal Graph Neural Network (MGNN) for predicting cancer survival, which utilizes multimodal data in a unified framework. In this framework, the problem of predicting cancer survival is solved as a learning task of patient classification.
- Extensive experiments are conducted to evaluate the performance of MGNN. We find MGNN can outperform start-of-art baselines on real-world datasets for patient classification task.

The rest of the paper is organized as following. In Section 2, we present the motivations. Related work is discussed in Section 3. Section 4 presents our proposed MGNN framework for patients representation learning. The experiments are presented in Section 5 and conclusion and discusion of the paper is presented in Section 6.

2 Motivations

In this section, we introduce the motivations for cancer prognosis prediction using multimodal data. The major challenges for cancer survival prediction are as follows.

(1) How to utilize the structure information between patients and multimodal medical data.

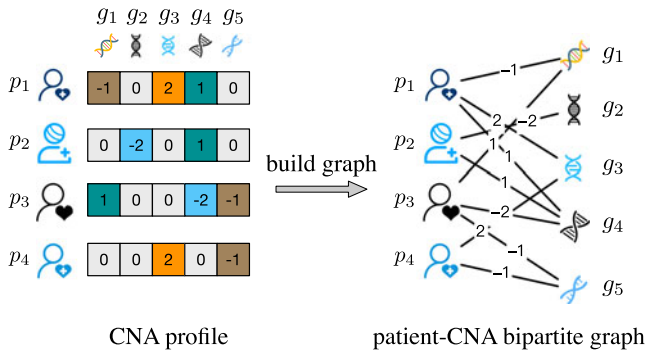

Fig. 1. An example of construction patient-CNA bipartite graph. In copy number alteration profile, each row represents the expression of all genes of the patient, and each column represents the expression of the gene of all patients. For example, the gene g_1, g_3 and g_4 of patient p_1 are not neutral/no change (0), patient p_1 will establish a edge with g_1, g_3 and g_4 respectively, and these edge weight are -1, 2 and 1.

In order to leverage the rich information in graph-structured data between patients and multimodal medical data, it is of great importance to learn effective node representation from the graph topological structure. One our intuition is that the node representation learning could be integrated by recursively aggregating and absorbing the continuous feature vectors from local neighborhoods. Motivated by this intuition, we aim to construct the bipartite graph using structure information between patients and multimodal medical data (gene expression profile or CNA profile) to learn effective node representation. Specifically, for patient gene expression profile whose data is expression intensity levels and we processed gene data into three categories: under-expression (-1), over-expression (1), baseline (0). If the gene is under-expression or over-expression, an edge will be established between the patient and the gene. Naturally we assign gene expression intensity level to edge weight. Thereafter we construct a patient-gene graph which is a bipartite graph. For patient CNA profile, we directly utilize the original data with five discrete values: homozygous deletion (-2); hemizygous deletion (-1); neutral/no change (0); gain (1); high level amplification(2). If the CNA is not neutral/no change (0), we will establish an edge between the patient and the CNA, so we can obtain a patient-gene bipartite graph with edge weights. An example of construction patient-CNA bipartite graph, as shown in Fig. 1.

One of the key component of graph neural networks is the feature aggregation function that aggregates and transforms the feature vectors from neighborhoods. After constructing the bipartite graph, we use graph neural network on the patient-gene and patient-CNA bipartite graph to obtain the embedding representation of the patients respectively, which means that it explores the inter-modality information of each modality data.

(2) How to utilize the features of medical data from different modalities.

It is far from enough to make classification for predicting the survival of cancer patients by using only single modality. Thus, we utilize multimodal features for predicting analysis because different modalities of data typically carry complement information of cancer patients. Multimodal data involves relating information from multiple sources including gene expression, CNA profile and clinical data.

The existing methods use feature extraction to make human breast cancer prognosis prediction which will lose the accuracy of patient embedding representation. To avoid this problem, we utilize the features of medical data from different modalities mainly in the following two aspects: intra-modality coordinated representations and inter-modality joint representations. For exploring features information between single intra-modality coordinated representations, we used all gene expression profile or CNA profile to construct the bipartite graph. Subsequently we used the graph neural network on the patient-gene bipartite graph to obtain an embedding representation of each patient and this is described as a patient embedding representation obtained using gene expression profile. Similarly, we can also use CNA data to obtain embedding representations of each patient. Another important process is to use the features information of inter-modality joint representations in medical data from different modalities, we concatenate above two embedding representations using gene expression profile or CNA profile and clinical data with standard normalization as the final embedding representation for each patient to do classification task.

These challenges motivated us to investigate a novel approach considering the task information. In the rest of the paper, we will present our method as well as its implementation.

3 Related Work

In this section, the related works, which are most close to our work, are briefly reviewed from the following three aspects: multimodal representation, graph embedding and graph neural networks.

3.1 Multimodal Representation

Cancer patients medical data is usually represented with different low-level features, and different types of cancer patients medical data, namely multimodal data, often coexist in many data sources. It is interesting and challenging to learn representations from multiple low-level features for multimodal representation. To learn feature representations from multiple aspects, multimodal representation has been successfully applied to many tasks, such as action or expression recognition [25], [26], [27], image or video classification [28], [29], medical diagnosis [30], [31], [32], disease diagnosis [33] and clinical prediction [34]. Most of the research conducted thus far on multimodal representation, either concatenates the single modality features and makes a classification task [35], or uses the classification results for each modality to improve the result by aggregation multimodal representation [22], [36]. In our work we use also both levels of fusion to obtain multimodal representation of cancer patients.

Regarding multimodal data representation, in [37], a unified framework for multimodal content retrieval is presented which supports retrieval of rich media objects as unified sets of different modalities, by efficiently combining all unimodal heterogeneous similarities to a global one according to an automatic weighting scheme. In [38], they present a multi-modality classification framework to efficiently exploit the complementarity in the multimodal data. Similarities from multiple modalities are then combined in a nonlinear graph fusion process, which generates a unified graph for final classification. [35] presents a multimodal framework for assessing the emotional and cognitive experience of blind and visually impaired people when navigating in unfamiliar indoor environments based on mobile monitoring and fusion of electroencephalography and electrodermal activity signals.

Multimodal representation has become a prominent concept in bioinformatics research, which reflects that cancer prognosis prediction is an amalgamation of various modality. Simple concatenation of multimodal medical data may make the parameter space complex due to the heterogeneity of multimodal data, and is thus bad in exploring the complementarity among data from different modalities. So it is necessary to develop an advanced method to extract deep information from the integrated multimodal data. To the best of our knowledge, we are the first to study cancer prognosis prediction using graph neural network on graph learning in a semisupervised setting and derive generalization model using multimodal medical data.

3.2 Graph Embedding

Graph-structured data has been extensively studied since the past decades for its ubiquity in real-world data and efficacy in fulfilling tasks, such as classification, clustering and recommendation. Meanwhile, graph embedding has emerged as an efficient and effective representation learning approach for graph-structured data. Although, graph embedding has recently become a paradigm to represent nodes by low-dimensional vectors, aiming to bridge the gap between graph analysis and machine learning techniques. Next, we briefly review some representative graph embedding methods. It is well recognized that graph data is sophisticated and challenging.

Early graph embedding methods, also called network embedding, are studied as a dimension reduction problem [39]. However, many methods focus on the pairwise similarity. How to preserve the high-order proximities becomes an attracting research problem very recently. For example, DeepWalk [40] uses local information obtained from truncated random walks to learn latent representations by treating walks as the equivalent of sentences. node2vec [41] extends this idea and propose a biased second order random walk model. LINE [42] optimizes an objective function which aims to preserve both the local and global network structures. M-NMF [43] is a unified framework, which enables the learned representations of nodes to preserve both of the microscopic and community structures. Furthermore, a semi-supervised deep model was proposed in SDNE [44], which has multiple layers of non-linear functions, thereby being able to capture the highly non-linear network structure. Besides, some recent works, such as [45] and [46], affiliate node attribute into the networks and smoothly embed both attribute information and topology structure into a low-dimensional representation.

With proper graph embedding, we can easily apply graph neural network to process biological networks data. Intuitively, it effectively learns effective node representation from the biological graph topological structure.

3.3 Graph Neural Networks

Over the past several years, the graph neural networks has achieved great success in many applications, such as

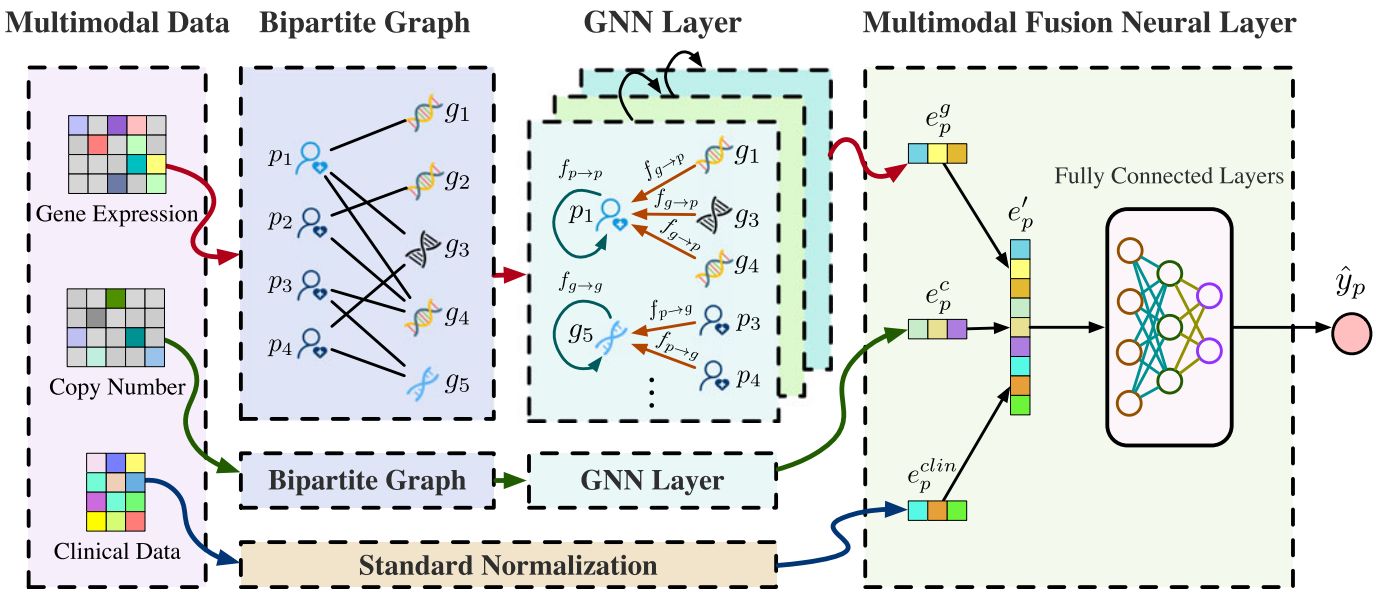

Fig. 2. The overall architecture of the proposed MGNN. There are three main parts in MGNN: bipartite graph, GNN layer and multimodal fusion neural layer. With the input multimodal data such as gene expression profile and CNA, (1) bipartite graph further expresses their potential relation (e.g, the edges between patient p_i and gene expression g_j); (2) the GNN layer, which aggregates information of neighbors in bipartite graph by GNN model; and (3) the multimodal fusion neural layer, which fuses the medical features of multimodal data. For the output labels $\widehat{y}_p$, the short term survivors are labeled as 0 and long term survivors are labeled as 1.

machine translation [47], object detection[48], [49], and recommendation algorithms[24], [50]. Graph neural networks are neural networks that learn node or graph representations from graph-structured data. Graph neural networks have emerged as a powerful neural architecture to learn embedding of nodes and have showed promising results on node classification.

Various graph neural networks have been proposed to encode the graph-structure information. Most of the existing graph neural networks learn the node representation by recursively aggregating the continuous feature vectors from local neighborhoods in an end-to-end fashion. Graph Convolutional Network (GCN) [51] defined a scalable approach for semi-supervised learning on graph-structured data that is based on an efficient variant of convolutional neural networks which operate directly on graphs. The convolution of GCN was a special form of Laplacian smoothing on graph, which explained the over-smoothing phenomena brought by many convolution layers. Graph attention networks (GAT) [52] proposed novel neural networks architectures that operate on graph-structured data, leveraging masked self-attentional layers to address the shortcomings of prior methods based on graph convolutions or their approximations. Even though, most of the graph neural networks architectures are motivated by the success of deep learning on grid-like data, graph neural networks have already been demonstrated useful on bioinformatics modeling [53], [54]. GRAM [53] represented a medical concept as a combination of its ancestors in the medical ontology using an attention mechanism. Another work worth mentioning is GAMENet [54], which also used graph neural networks on drug-drug-interaction graphs to assist the medication recommendation task.

Finally, after reviewing multimodal representation, graph embedding and graph neural networks, we briefly review existing works of human breast cancer prognosis prediction. Since deep learning method yielded great success in many fields, some researchers also have tried to explore different neural network structures for human breast cancer prognosis prediction task. For instance, the authors in [22] propose a multimodal deep neural network for the prognosis prediction of breast cancer. This method achieves a better performance than the prediction methods with single-dimensional data. Some methods rely on selected gene expression or CNA profile to create a predictive model. Xiaoyi *et al.* [23] adopted the support vector machine-based recursive feature elimination approach for gene selection and prognosis prediction. Cuong *et al.* [55] uses a machine learning method based on random forest classifier and feature selection technique on diagnosing and prognosticating breast cancer. The authors in [56] used the logistic regression method to make prognostic predictions. However, existing neural methods [22], [23], [55], [56] mainly rely on selected gene expression data or CNA data to create a predictive model, which may lose effective information. Based on this, obtaining embedding representation of each patient using a graph neural network is one of the main concerns.

Previous work has not explicitly considered graph structures information and multimodal medical data for node representation learning of predicting the survival of cancer patients. Our proposed method MGNN is designed to address this issue by compressing the learned patient representation according to graph structures and multimodal medical data.

4 Methodology

In this section, we first define the problem formulation. Then, the detailed design of MGNN, which learns classification information from multimodal data, is shown in Fig. 2. The overall structure is composed of four parts: multimodal data, the bipartite graph, the GNN layer, and the multimodal fusion neural layer.

4.1 Problem Formulation

In this study, multimodal data are composed of the gene expression profile data, the CNA profile data and the clinical data. It is expressed as follows:

$$\mathcal{X}_p = \{X_g, X_c, X_{clin}\} \in \mathbb{R}^{N\times(m+n+k)}, \tag{1}$$

where $X_g \in \mathbb{R}^{N\times m}$, $X_c \in \mathbb{R}^{N\times n}$ and $X_{clin} \in \mathbb{R}^{N\times k}$ stand for the gene expression profile, the CNA profile and the clinical data respectively. m, n and k represent the dimension of the gene data, the CNA data and the clinical data respectively and N is the number of patients. We first build bipartite graphs via $X_g \in \mathbb{R}^{N\times m}$ and $X_c \in \mathbb{R}^{N\times n}$, and then the node of bipartite graph aggregates information of neighbors by GNN model. Specifically, the final embedding of the patient can be shown as below:

$$\mathcal{E}_p = \{E_g, E_c, E_{clin}\} \in \mathbb{R}^{N\times(d_1+d_2+k)}, \tag{2}$$

where $E_g \in \mathbb{R}^{N\times d_1}$, $E_c \in \mathbb{R}^{N\times d_2}$ and $E_{clin} \in \mathbb{R}^{N\times k}$ respectively represent the patient embedding from multimodal data, d_1 and d_2 represent the size of the embedding dimension. $e'_p \in \mathcal{E}_p$ denotes the multimodal representation vector for each patient.

The goal of the study is to distinguish short term survivors and long term survivors (breast cancer usually adopt less 5-year survival), and the formal definition of the problem is as follows:

Input: multimodal data $\mathcal{X}_p = \{X_g, X_c, X_{clin}\}$.

Output: the predictive classification of short term survivors or long term survivors.

In order to predict the survival of cancer patients from multimodal data, we design a novel Multimodal Graph Neural Network (MGNN) as illustrated in Fig. 2.

4.2 MGNN

In this subsection, we introduce the detailed design of the MGNN for predicting the survival of cancer patients. The general framework of the proposed method is shown in Fig. 2, from which we can see that there are three main processing, one for each modality. Our method aim to utilize the inter-modality and intra-modality correlation to learn the final representations of patients.

Multimodal Data. In our work, multimodal data exist in many data sources, such as clinical data, gene expression and CNA profile. Intuitively, to learn representation from multimodal data instead of single modality data is beneficial for embedding of patients. It is an challenging problem to learn representation of patients from multimodal data by effective fusion of multiple independent features. First, we explore structural information from each feature type for intra-modality representation. Second, we further explore inter-modality canonical correlation via multimodal fusion of clinical data and structural information.

Bipartite Graph. To process graph data effectively, the first critical challenge is graph-structured data representation, that is, how to represent graphs properly so that advanced prediction tasks can be conducted efficiently in both time and space. In order to establish the link between patients and multimodal data, we utilize the gene expression and the CNA profile which belong to the multimodal data to construct bipartite graph. According to the previous method [57], the gene expression features are normalized and processed into three categories: under-expression (-1), over-expression (1), baseline (0). For each patient, an edge will be built between the patient and the gene, only if the gene is not properly expressed (under-expression or over-expression). The edge weight between the patient and the gene indicates the gene expression intensity level. Finally we construct a patient-gene bipartite graph with edge weights. Obviously, we can intuitively understand the gene expression data affecting patients from the patient-gene bipartite graph.

For CNA data, we directly utilize the original data with five discrete values: homozygous deletion (-2); hemizygous deletion (-1); neutral/no change (0); gain (1); high level amplification (2). A patient-CNA bipartite graph is constructed by using the similar approach of the construction of patient-gene bipartite.

GNN Layer. After getting patient-gene bipartite graph, the initial representation matrix of patient-gene bipartite E is as follows:

$$E = [\underbrace{e_{p_1}^{(0)}, e_{p_2}^{(0)}, \ldots, e_{p_N}^{(0)}}_{patients\ embedding}, \underbrace{e_{g_1}^{(0)}, e_{g_2}^{(0)}, \ldots, e_{g_M}^{(0)}}_{genes\ embedding}], \tag{3}$$

where N and M stand for the number of patient and gene, respectively, $e_p^{(0)}$, $e_g^{(0)}$ are served as the initialization of patient embedding and gene embedding respectively.

In order to get the genetic information of patients more efficiently, we aggregate the messages propagated from the neighborhood of p to refine the embedding of p. Specifically, we define the aggregation function as below:

$$e_p^{(l+1)} = \sigma(f_{p\to p}(e_p^{(l)}) + \sum_{g\in\mathcal{N}_p} f_{g\to p}(e_p^{(l)}, e_g^{(l)})), \tag{4}$$

where $e_p^{(l+1)}$ denotes the embedding of patient p obtained at the $(l+1)$th GNN layer, $\mathcal{N}_p$ is neighbor set of patient p, $\sigma(\cdot)$ denotes activation function $LeakyReLU$, and $f(\cdot)$ is the representation encoding function that can be shown as below:

$$\begin{cases} f_{p\to p}(e_p^{(l)}) = W_1^{(l)} e_p^{(l)} \\ f_{g\to p}(e_p^{(l)}, e_g^{(l)}) = \alpha_{gp}(W_1^{(l)} e_g^{(l)} + W_2^{(l)}(e_g^{(l)} \odot e_p^{(l)})), \end{cases} \tag{5}$$

where $W_1^{(l)}, W_2^{(l)} \in \mathbb{R}^{d_l\times d_{l+1}}$ are the trainable weight matrices, α_{gp} is edge weight between patient p and gene g, and $\odot$ denotes the element-wise product. $e_p^{(l)}$ and $e_g^{(l)}$ are the representation of patients and genes in l layer of GNN. An example of the node feature aggregation on GNN Layer, as shown in Fig. 3.

The embedding of patients based on the patient-gene bipartite graph, through the GNN layer, is e_p^g. Similarly, the embedding of patients e_p^c can be obtained by patient-CNA bipartite graph.

Standard Normalization. Clinical data is a kind of sequence data that contains different clinical characteristics, such as sex, mutation count, invasive carcinoma diagnosis age and stage at diagnosis etc. Since different patients have different clinical manifestations, obviously, the corresponding clinical data are also different. We perform min-max normalization

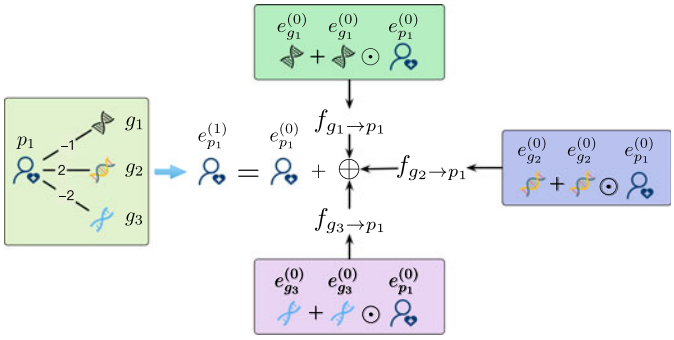

Fig. 3. An example of the node feature aggregation on GNN Layer. $e_{p_1}^{(0)}$, $e_{g_1}^{(0)}$, $e_{g_2}^{(0)}$, $e_{g_3}^{(0)}$ are served as the initialization of p_1 embedding and g_1, g_2, g_3 embedding respectively.

operations on the patient's clinical data as an embedding representation of the patient. Min-Max normalization is formally defined by:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}. \tag{6}$$

Where X_{max} and X_{min} respectively denote the maximum and minimum values of each column of clinical data. The clinical data after min-max normalization is denoted as e_p^{clin}.

Multimodal Fusion Neural Layer. The multimodal fusion neural layer is designed to concatenate the different modalities, so that their complement information can be captured in the embedded feature space. Specifically, after getting the representation e_p^g, e_p^c and e_p^{clin} for patient p by multimodal data, these embedding are linked together as the final multimodal embedding e'_p

$$e'_p = e_p^g || e_p^c || e_p^{clin}. \tag{7}$$

Then, the multimodal fusion embedding e'_p with multiple fully connected layers is used to predict the survival of cancer patients as follows:

$$\hat{y}_p = softmax(\rho(W_3^{(l)} e_p'^{(l)} + b^{(l)})), \tag{8}$$

where $W_3^{(l)}$ is the trainable weight matrix, ρ denotes activation function $tanh$, $e_p'^{(l)}$ denotes the final multimodal representation for patient p at the layer l, and $b^{(l)}$ is the bias vector. Finally we use softmax function and obtain the final prediction score $\hat{y}_p$.

4.3 Optimization

For model optimization, our method could be trained with supervised setting. Besides, we use L2 regularization to prevent over-fitting of our model. Based on the cross-entropy loss, the objective function could be defined as follows:

$$\begin{aligned} L(y_p, \hat{y}_p) = &-\frac{1}{N}\sum_{i=0}^{N}[y_{p_i} log \hat{y}_{p_i} - (1 - y_{p_i}) log(1 - \hat{y}_{p_i})] \\ &+ \frac{1}{\lambda}\sum_{l=1}^{L}\sum_{t=1}^{T}\sum_{i=1}^{d_l}\sum_{j=1}^{d_{l+1}} {w_{tij}^l}^2, \end{aligned} \tag{9}$$

where y_{p_i} is the actual label, $\hat{y}_{p_i}$ is the predictive scores and N is the batch size, L and T is the number of the embedding layers and trainable weight matrices W, respectively. $W_t^l = \{w_{tij}^l\}_{d_l \times d_{l+1}}$ is the lth weight matrix.

TABLE 1
The Properties of the Breast Cancer Dataset

Data Category	Feature Number
Clinical	28
Gene Expression	24369
Copy Number	22544

5 Experiments

In this section, we first introduce the experimental datasets, compared methods, evaluation metrics and parameter setting. Then, we show that our method consistently outperforms the state-of-the-art baselines in terms of cancer survival prediction accuracy. Ablation study and parameter sensitivity analysis are also conducted to demonstrate the effectiveness of our major model components. the discriminators. We also investigate the effects of using other four different cancer datasets and present a robustness verification with our method.

5.1 Experimental Settings

Datasets. To demonstrate the effectiveness of MGNN,[1] we perform exhaustive experiments on a publicly available benchmark datasets. The datasets contain multimodal data about more than 2,500 breast cancer patients, including gene expression profile, copy number alteration (CNA) profile and clinical data. In particular, the gene expression profile includes approximately 24369 genes and CNA profile contains approximately 22544 genes in the breast cancer dataset. The detailed properties of the dataset are shown in Table 1. According to [57], gene expression profile is processed into three categories: under-expression (-1), over-expression (1), baseline (0). For CNA data, we directly utilize the original data with five discrete values: homozygous deletion (-2); hemizygous deletion (-1); neutral/no change (0); gain (1); high level amplification(2). Each patient has 35 clinical features, including sex, mutation count, invasive carcinoma diagnosis age and stage at diagnosis etc. Finally, we utilize 28 clinical features in this experiment. The summary statistics of the dataset is shown in Table 3.

To comprehensively evaluate our proposed method, we adopt ten-fold cross validation in our experiments. Specifically, we randomly divide all patients into ten subsets. For each round of training, each subset will be used as a test set, and the remaining nine subsets are divided into a training set (80 percent) and a validation set (20 percent). The prediction score is the average of the output of ten rounds.

Compared Methods. We compared MGNN with four strong state-of-the-art models in the human breast cancer prognosis prediction task. The baselines of this work are as follows:

1. Our supplemental material are available at https://github.com/tflyu/mgnn.

TABLE 2
The Performance Comparison of the State-of-the-Art Approaches on the Breast Cancer Dataset

Methods	Acc	Pre	Sn	MCC	AUC
LR	0.760	0.549	0.183	0.209	0.663
RF	0.791	0.766	0.226	0.337	0.801
SVM	0.805	0.708	0.365	0.407	0.810
MDNNMD	0.826	0.749	0.450	0.486	0.845
MGNN	**0.954**	**0.964**	**0.976**	**0.875**	**0.983**

- *MDNNMD*: In this work, a novel multimodal deep neural network by integrating multi-dimensional data is proposed for the prognosis prediction of breast cancer [22]. The novelty of the method lies in the design of their method's architecture and the fusion of multi-dimensional data including clinical data, gene expression, and CNA profile.
- *SVM*: This model is proposed with an efficient feature selection method based on Support Vector Machine (SVM) for breast cancer prognosis prediction [23].
- *RF*: This method for breast cancer prognosis prediction based on Random Forest (RF) combined with feature selection and achieved the highest classification accuracy [55].
- *LR*: In this study, the model used Logistic Regression (LR) for predictive analysis of breast cancer patients [56].

Evaluation Metrics. In our experiments, we evaluate the MGNN with five metrics including the Area Under the Curve (AUC) of the receiver operating characteristics (ROC) curve, Sensitivity (Sn), Accuracy (Acc), Precision (Pre) and Matthew's Correlation Coefficient (MCC). In these evaluation metric, TP, FP, TN and FN stand for true positive, false positive, true negative and false negative, respectively. T is the total number of positives; N is the total number of negatives. We introduce them as follows:

- *Sensitivity (Sn)*: Represents the proportion of all positive samples that are correctly classified, and measures the classifier's ability to recognize positive samples. The definition is shown below:

$$Sn = \frac{TP}{TP + FN}. \tag{10}$$

- *Specificity (Sp)*: Defined as the proportion of samples that are actually negative, classified as negative. It can be defined as

$$Sp = \frac{TN}{TN + FP}. \tag{11}$$

- *Accuracy (Acc)*: Defined as the proportion of correctly classified samples to the total number of samples. Are defined as follows:

$$Acc = \frac{TP + TN}{TP + TN + FN + FP}. \tag{12}$$

- *Precision (Pre)*: Defined as the proportion of samples that were actually positive, which was classified as positive. The definition is shown below:

$$Pre = \frac{TP}{TP + FP}. \tag{13}$$

- *Matthew's Correlation Coefficient (MCC)*: The Matthews correlation coefficient is the most informative single fraction for establishing a binary classifier to predict mass in a confounding matrix environment. Are defined as follows:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{T \times (TP + FP) \times F \times (TN + FP)}}. \tag{14}$$

Parameter Settings. The proposed method is implemented with Tensorflow and optimized by Adam with a learning rate of 0.001. The parameters in Section 4 set as $m = 256$, $n = 256$, $k = 28$, and $d_1 = d_2 = 128$. The embedding $e_g^{(0)}$ and $e_c^{(0)}$ are 64 dimensional and initialized randomly. After obtaining the embedding e_p', it is filled into the full connection layers with 3 layers, and each hidden layer contains 1000, 500 and 100 units.

5.2 Results and Analysis

We analyze the experimental results on breast cancer dataset to verify the effectiveness of MGNN and we also design ablation tests to emphasize that multimodal representation can improve the accuracy of cancer prognosis prediction. In addition, we conduct additional experiments to investigate the sensitivity of MGNN with respect to this parameters. Finally, we use other cancer dataset to verify the robustness of the proposed MGNN.

Performance Comparison. We compare the results of our method for breast cancer survival prediction with that of four state-of-the-art models including MDNNMD[22], SVM [23], RF[55] and LR[56]. We directly use the experimental results of MDNNMD. Table 2 shows the experimental results of these models with the five evaluation metrics including Acc, Pre, Sn, MCC and AUC. Compared to these baselines, it can be seen that our proposed MGNN achieves the best performance among all contrastive methods. For example, the accuracy of our method reaches 95.4 percent,

TABLE 3
Summary Statistics of Breast Cancer Dataset

Category	Number
Cut-off(years)	5
Total population	1903
Long time survivors	1432
Short time survivors	471

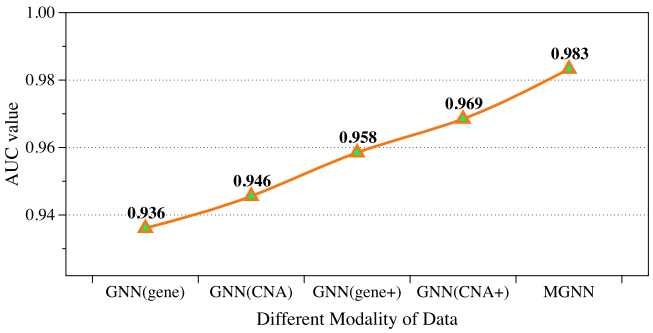

Fig. 4. The AUC value of model variant using different modality of data.

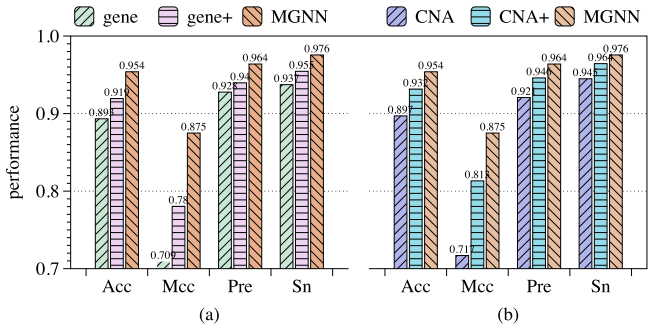

Fig. 5. Results of the ablation experiments with the relative performances compared with complete MGNN in all the metrics.

which is higher 12.8 percent than the second high accuracy of MDNNMD method. Due to the expressive power and flexibility of graph-structured data, MGNN has achieved the best performance, which tries to obtain the effective representation of cancer patients. By leveraging the node adjacency matrix, these GNN models analogously define convolution operators on graphs and have obtained promising performance in tasks like node classification. It shows the power of MGNN method to extract multimodal data features. Besides, graph neural networks can effectively assemble bipartite graph structure interaction information for predicting the survival of cancer patients.

5.3 Ablation Test

To further explore the effect of multimodal fusion in our proposed model, we adopt graph neural network method on different modality data for predicting the survival of cancer patients. we compare MGNN with GNN(gene), GNN (gene+), GNN(CNA) and GNN(CNA+), where GNN (gene), GNN (CNA) are parts of our model and described as only using gene expression or CNA profile to construct bipartite graph for cancer survival prediction, respectively. The meanings of GNN(gene+) and GNN(CNA+) are to use two modality data such as gene expression and clinical or CNA profile and clinical. For example, GNN(gene+) means using gene expression to construct bipartite graph and then obtain the patients representation. Subsequently, we concatenate patients representation and clinical data with standard normalization as the final embedding representation for each patient to do classification task. GNN(CNA+) is constructed by using the similar approach of GNN(gene+).

As shown in Fig. 5, all the evaluation metrics of MGNN achieve the best performance, which shows that MGNN can effectively fuse multimodal data from different dimensions. For example, the MGNN obtains the highest Acc value (0.954), while Acc value of GNN(gene), GNN(CNA), GNN (gene+) and GNN(CNA+) are 0.893, 0.897, 0.919 and 0.932. Meanwhile, the MCC value of MGNN is 0.875, which is 16.6, 15.8, 9.5 and 6.2 percent higher than GNN(gene), GNN (CNA), GNN(gene+) and GNN(CNA+). The AUC value for each method is obtained and shown in Fig. 4, MGNN achieves better overall performance than those methods using single or two modality data.

In Fig. 5, it is shown the results of the ablation test. MGNN using multimodal data demonstrated high accuracy compared to single modality approaches (GNN(gene) and GNN(CNA)). The results demonstrate that the estimation performance of MGNN is the highest for the evaluation metrics. The difference in the accuracy of estimation is related to different modality data performance. Thus, this result indicates that the combination of clinical data, gene expression and copy number profile improve the accuracy of estimation by complementarity explanation. However, the result confirmed the effectiveness of multimodal graph neural network in predicting the survival of cancer patients. Moreover, each metric score indicates higher performance of the proposed methods MGNN. The results support the effectiveness of MGNN on multimodal graph neural networks to improve the accuracy of estimation in both GNN (gene), a single modality data, and GNN(gene+), two modality data, of the multimodal graph neural network.

5.4 Parameter Sensitivity Analysis

After investigating the risk of ablation test, we point out the parameter sensitivity analysis of the training strategy in terms of training performance. In this work, different number of hidden layers is the most important parameters which may affect the performance of MGNN. In order to explore the effects of setting different number of hidden layers on the results of predicting the survival of cancer patients with same dataset, we are following the experimental protocols as in the breast cancer experiment. As shown in Fig. 6, when utilizing patient-gene bipartite graph or patient-CNA bipartite graph to obtain the embedding of each patient, we studied the effects of different number of hidden layers on classification results. For example, the corresponding Acc values obtained by 1-hidden layers, 2-hidden layers and 3-hidden layers using patient-gene bipartite

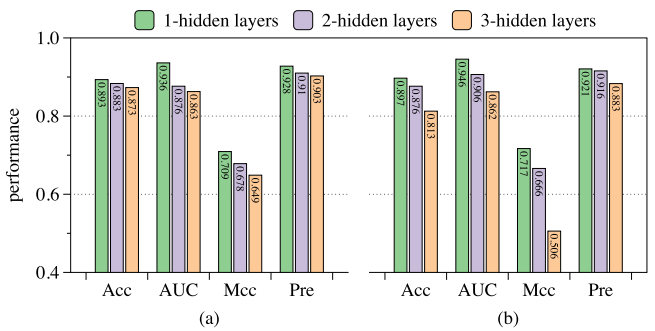

Fig. 6. Parameter sensitivity analysis of different hidden layers, when using gene or CNA to obtain the embedding representation of each patient. (a) utilizing gene expression profile. (b) utilizing CNA profile.

TABLE 4
The Properties of the Robustness Verification Datasets

Datasets	Data Category	Feature Number
BIC	Clinical	36
	Gene Expression	20530
	Copy Number	25128
MeCC	Clinical	34
	Copy Number	474
PALL	Copy Number	20641
PLC	Copy Number	24203

TABLE 5
The Overall Information of Robustness Verification Datasets

Category	Number			
	BIC	MeCC	PALL	PLC
Cut-off	1(years)	18(months)	4(years)	3(months)
Total population	1051	1010	583	917
Long time	888	731	464	656
Short time	163	279	119	261

graph are 0.893, 0.883 and 0.873, respectively. Meanwhile the corresponding MCC values obtained by 1-hidden layers, 2-hidden layers and 3-hidden layers using patient-CNA bipartite graph are 0.717, 0.666 and 0.506, respectively. As shown in Fig. 6, the results showed that the performance were the best when only one hidden layer was used. This proves that it is the most accurate to directly use the abnormally expressed genes of cancer patients to obtain patients representation.

5.5 Robustness Verification

Robustness is crucial for deep learning model. To verify the robustness of the proposed MGNN, we use other cancer datasets to verify the robustness of this modal including Breast Invasive Carcinoma (BIC), Metastatic Colorectal Cancer (MeCC), Pediatric Acute Lymphoid Leukemia - Phase II (PALL) and Pan-Lung Cancer (PLC). The detailed properties of the robustness verification datasets are shown in Table 4. The format of these cancer datasets are same as the breast cancer dataset. For example, PALL dataset contains CAN profile of 583 cancer patients. The CNA profile data of PALL patients includes approximately 20641 genes, and we adopt only CNA data to construct a bipartite graph, and then use MGNN to obtain the embedding representation of each PALL patient on the bipartite graph and finally make classification predictions. Similarly, other datasets are processed using the same operation to make classification predictions. The overall information of the robustness verification datasets is shown in Table 5.

As shown in Table 6, we use MGNN to calculate the overall performance results of different datasets. It can be seen from the results that MGNN shows good performance on different cancer datasets. For example, the MGNN obtains highest AUC value (0.959) of BIC dataset, while AUC value of PLC dataset, MeCC dataset and PALL dataset are 0.949, 0.896 and 0.722. Meanwhile, the MGNN obtains the highest Acc value (0.897) of PLC dataset, while Acc value of BIC dataset, MeCC dataset and PALL dataset are 0.866, 0.826 and 0.822. Table 6 shows that the proposed MGNN model using the features of multimodal data on bipartite graphs consistently yield new state-of-the-art performance in predicting the survival of cancer patients tasks on different cancer datasets. These results have demonstrated the effectiveness of applying graph neural networks operations on patients bipartite graph data.

TABLE 6
MGNN Prediction Performance on Other Cancer Dataset for Robustness Verification

Datasets	Acc	Pre	Sn	MCC	AUC
BIC	0.866	0.873	0.887	0.730	0.959
PLC	0.897	0.912	0.947	0.741	0.949
MeCC	0.826	0.882	0.880	0.564	0.896
PALL	0.822	0.862	0.928	0.350	0.722

6 Conclusion and Discusion

In this paper, inspired by the great success of graph neural network in grid-structured data, we propose a novel method MGNN for the cancer survival prediction. MGNN learns the patient representation by recursively aggregating the continuous feature vectors from local neighborhoods and utilizing the inter-modality and intra-modality correlation of multimodal data with a unified framework. The multimodal data includes clinical data, gene expression profile and CNA profile. We explore structural information from each feature type for intra-modality learning, and we further explore inter-modality canonical correlation via multimodal fusion of clinical data and structural information. The experimental results showed the consistent performance improvement by the proposed method over the state-of-art baseline methods on real-world breast cancer dataset. Our proposed method is highly scalable, which is demonstrated via evaluation on other cancer datasets including breast invasive carcinoma, metastatic colorectal cancer, pediatric acute lymphoid leukemia - Phase II and pan-lung cancer.

Future research should consider the potential effects of multimodal data more carefully. For example, directions for future work include using more multimodality data (MRI, CT, etc) to learn the patient representation. It would also be of interest to extend the experiments to other datasets and explore the structure information between patients and multimodal medical data.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61873288, U1972423), Zhejiang Lab (2019KE0 AB01), Hunan Provincial Innovation Foundation for Postgraduate (CX20200108) and the Fundamental Research Funds for the Central Universities of Central South University (2021zzts0740).

References

[1] Z. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec, "GNN-Explainer: Generating explanations for graph neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 9244–9255.

[2] J. Gao, T. Lyu, F. Xiong, J. Wang, W. Ke, and Z. Li, "MGNN: A multimodal graph neural network for predicting the survival of cancer patients," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2020, pp. 1697–1700.
[3] W. Chen *et al.*, "Semi-supervised user profiling with heterogeneous graph attention networks," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, 2019, pp. 2116–2122.
[4] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang, "LightGCN: Simplifying and powering graph convolution network for recommendation," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2020, pp. 639–648.
[5] C. Yan, G. Duan, Y. Zhang, F.-X. Wu, Y. Pan, and J. Wang, "Predicting drug-drug interactions based on integrated similarity and semi-supervised learning," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, early access, Apr. 16, 2020, doi: 10.1109/TCBB.2020.2988018.
[6] C. Yan, G. Duan, Y. Pan, F. Wu, and J. Wang, "DDIGIP: Predicting drug-drug interactions based on Gaussian interaction profile kernels," *BMC Bioinf.*, vol. 20, no. 15, pp. 1–10, 2019.
[7] M. Zeng, F. Zhang, F. Wu, Y. Li, J. Wang, and M. Li, "Protein–protein interaction site prediction through combining local and global features with deep neural networks," *Bioinformatics*, vol. 36, no. 4, pp. 1114–1120, 2020.
[8] M. Milano, T. Milenković, M. Cannataro, and P. H. Guzzi, "L-HetNetAligner: A novel algorithm for local alignment of heterogeneous biological networks," *Sci. Rep.*, vol. 10, 2020, Art. no. 3901.
[9] X. Wang, R. Wang, Z. Bao, J. Liang, and W. Lu, "Effective medical archives processing using knowledge graphs," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2019, pp. 1141–1144.
[10] X. Lin, Z. Quan, Z.-J. Wang, T. Ma, and X. Zeng, "KGNN: Knowledge graph neural network for drug-drug interaction prediction," in *Proc. 29th Int. Joint Conf. Artif. Intell.*, 2020, pp. 2739–2745.
[11] T. Ma and A. Zhang, "AffinityNet: Semi-supervised few-shot learning for disease type prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 1069–1076.
[12] Y. Wu, Y. Yang, H. Nishiura, and M. Saitoh, "Deep learning for epidemiological predictions," in *Proc. 41st Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2018, pp. 1085–1088.
[13] Y. Xiao, J. Wu, Z. Lin, and X. Zhao, "A deep learning-based multi-model ensemble method for cancer prediction," *Comput. Methods Programs Biomed.*, vol. 153, pp. 1–9, 2018.
[14] A. Cheerla and O. Gevaert, "Deep learning with multimodal representation for pancancer prognosis prediction," *Bioinformatics*, vol. 35, no. 14, pp. i446–i454, 2019.
[15] D. W. Kim, S. Lee, S. Kwon, and W. Nam, "Deep learning-based survival prediction of oral cancer patients," *Sci. Rep.*, vol. 9, 2019, Art. no. 6994.
[16] M. S. Bal, V. K. Bodal, J. Kaur, M. Kaur, and S. Sharma, "Patterns of cancer: A study of 500 Punjabi patients," *Asian Pacific J. Cancer Prevention*, vol. 16, no. 12, pp. 5107–5110, 2015.
[17] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Comput. Structural Biotechnol. J.*, vol. 13, pp. 8–17, 2015.
[18] P. Han *et al.*, "GCN-MF: Disease-gene association identification by graph convolutional networks and matrix factorization," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 705–713.
[19] F. Zhang *et al.*, "A deep learning framework for gene ontology annotations with sequence - and network-based information," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, early access, Jan. 23, 2020, doi: 10.1109/TCBB.2020.2968882.
[20] L. R. Martin, S. L. Williams, K. B. Haskard, and M. R. DiMatteo, "The challenge of patient adherence," *Ther. Clin. Risk Manage.*, vol. 1, no. 3, pp. 189–199, 2005.
[21] S. Lee, S. Lim, T. Lee, I. Sung, and S. Kim, "Cancer subtype classification and modeling by pathway attention and propagation," *Bioinformatics*, vol. 36, no. 12, pp. 3818–3824, 2020.
[22] D. Sun, M. Wang, and A. Li, "A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 16, no. 3, pp. 841–850, May/Jun. 2019.
[23] X. Xu, Y. Zhang, L. Zou, M. Wang, and A. Li, "A gene signature for breast cancer prognosis using support vector machine," in *Proc. 5th Int. Conf. BioMedical Eng. Informat.*, 2012, pp. 928–931.
[24] X. Wang, X. He, M. Wang, F. Feng, and T.-S. Chua, "Neural graph collaborative filtering," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2019, pp. 165–174.
[25] H. Li, J. Sun, Z. Xu, and L. Chen, "Multimodal 2D+ 3D facial expression recognition with deep fusion convolutional neural network," *IEEE Trans. Multimedia*, vol. 19, no. 12, pp. 2816–2831, Dec. 2017.
[26] Z. Liu, L. Zhang, Q. Liu, Y. Yin, L. Cheng, and R. Zimmermann, "Fusion of magnetic and visual sensors for indoor localization: Infrastructure-free and more effective," *IEEE Trans. Multimedia*, vol. 19, no. 4, pp. 874–888, Apr. 2017.
[27] Z. Qiao, X. Wu, S. Ge, and W. Fan, "MNN: Multimodal attentional neural networks for diagnosis prediction," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, 2019, pp. 5937–5943.
[28] M. Guillaumin, J. Verbeek, and C. Schmid, "Multimodal semi-supervised learning for image classification," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 902–909.
[29] J. Geng, Z. Miao, and X.-P. Zhang, "Efficient heuristic methods for multimodal fusion and concept fusion in video concept detection," *IEEE Trans. Multimedia*, vol. 17, no. 4, pp. 498–511, Apr. 2015.
[30] L. Xu, X. Wu, K. Chen, and L. Yao, "Multi-modality sparse representation-based classification for Alzheimer's disease and mild cognitive impairment," *Comput. Methods Programs Biomed.*, vol. 122, no. 2, pp. 182–190, 2015.
[31] E. A. Bernal *et al.*, "Deep temporal multimodal fusion for medical procedure monitoring using wearable sensors," *IEEE Trans. Multimedia*, vol. 20, no. 1, pp. 107–118, Jan. 2018.
[32] C. Tan, F. Sun, W. Zhang, J. Chen, and C. Liu, "Multimodal classification with deep convolutional-recurrent neural networks for electroencephalography," in *Proc. Int. Conf. Neural Inf. Process.*, 2017, pp. 767–776.
[33] F. Ma, Q. You, H. Xiao, R. Chitta, J. Zhou, and J. Gao, "KAME: Knowledge-based attention model for diagnosis prediction in healthcare," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, 2018, pp. 743–752.
[34] Y. Xu, S. Biswal, S. R. Deshpande, K. O. Maher, and J. Sun, "RAIM: Recurrent attentive and intensive model of multimodal patient monitoring data," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2018, pp. 2565–2573.
[35] K. Kalimeri and C. Saitis, "Exploring multimodal biosignal features for stress detection during indoor mobility," in *Proc. 18th ACM Int. Conf. Multimodal Interact.*, 2016, pp. 53–60.
[36] S. S. Mukherjee and N. M. Robertson, "Deep head pose: Gaze-direction estimation in multimodal video," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 2094–2107, Nov. 2015.
[37] D. Rafailidis, S. Manolopoulou, and P. Daras, "A unified framework for multimodal retrieval," *Pattern Recognit.*, vol. 46, no. 12, pp. 3358–3370, 2013.
[38] T. Tong *et al.*, "Multi-modal classification of Alzheimer's disease using nonlinear graph fusion," *Pattern Recognit.*, vol. 63, pp. 171–181, 2017.
[39] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.
[40] B. Perozzi, R. Al-Rfou, and S. Skiena, "DeepWalk: Online learning of social representations," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2014, pp. 701–710.
[41] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 855–864.
[42] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, "LINE: Large-scale information network embedding," in *Proc. 24th Int. Conf. World Wide Web*, 2015, pp. 1067–1077.
[43] X. Wang, P. Cui, J. Wang, J. Pei, W. Zhu, and S. Yang, "Community preserving network embedding," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 203–209.
[44] D. Wang, P. Cui, and W. Zhu, "Structural deep network embedding," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 1225–1234.
[45] X. Huang, J. Li, and X. Hu, "Label informed attributed network embedding," in *Proc. 10th ACM Int. Conf. Web Search Data Mining*, 2017, pp. 731–739.
[46] J. Li, H. Dani, X. Hu, J. Tang, Y. Chang, and H. Liu, "Attributed network embedding for learning in a dynamic environment," in *Proc. ACM Conf. Inf. Knowl. Manage.*, 2017, pp. 387–396.
[47] M. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," in *Proc. Conf. Empir. Methods Natural Lang. Process.*, 2015, pp. 1412–1421.

[48] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.
[49] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
[50] M. Nassar, "Hierarchical bipartite graph convolution networks," 2018, *arXiv:1812.03813*.
[51] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Representations*, 2017, pp. 1–14.
[52] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," in *Proc. Int. Conf. Learn. Representations*, 2018, pp. 1–12.
[53] E. Choi, M. T. Bahadori, L. Song, W. F. Stewart, and J. Sun, "GRAM: Graph-based attention model for healthcare representation learning," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2017, pp. 787–795.
[54] J. Shang, C. Xiao, T. Ma, H. Li, and J. Sun, "GAMENet: Graph augmented memory networks for recommending medication combination," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 1126–1133.
[55] C. Nguyen, Y. Wang, and H. N. Nguyen, "Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic," *J. Biomed. Sci. Eng.*, vol. 6, no. 5, pp. 551–560, 2013.
[56] M. F. Jefferson, N. Pendleton, S. B. Lucas, and M. A. Horan, "Comparison of a genetic algorithm neural network with logistic regression for predicting outcome after surgery for patients with nonsmall cell lung carcinoma," *Cancer: Interdisciplinary Int. J. Amer. Cancer Soc.*, vol. 79, no. 7, pp. 1338–1342, 1997.
[57] P. Stone and S. Lund, "Predicting prognosis in patients with advanced cancer," *Ann. Oncol.*, vol. 18, no. 6, pp. 971–976, 2007.

Jianliang Gao received the PhD degree from the Institute of Computing Technology, Chinese Academy of Sciences, Guangzhou, China. He is currently a professor with the School of Computer Science and Engineering, Central South University, China. His main research interests include bioinformatics, machine learning, and graph data mining.

Tengfei Lyu received the BS degree from Bohai University, Jinzhou, Liaoning, China, in 2019. He is currently working toward the graduate degree in the School of Computer Science and Engineering, Central South University, Changsha, China. His research interests include bioinformatics, machine learning, and graph neural network.

Fan Xiong received the BS degree from Central South University, Changsha, China, in 2018. He is currently working toward the graduate degree in the School of Computer Science and Engineering, Central South University, Changsha, China. His research interests include graph neural network and knowledge graph.

Jianxin Wang (Senior Member, IEEE) received the BEng and MEng degrees in computer engineering and the PhD degree in computer science from Central South University, Changsha, Hunan, China, in 1992, 1996, and 2001, respectively. He is currently the chair and a professor with the School of Computer Science and Engineering, Central South University. His current research interests include algorithm analysis and optimization, parameterized algorithm, bioinformatics, and computer networks.

Weimao Ke received the PhD degree from the University of North Carolina at Chapel Hill, Chapel Hill, North Carolina. He is currently an associate professor with Drexel University. His research is centered on information retrieval (IR), particularly the investigation of intelligent systems that support better connection and interaction between people and information. His recent focus is on decentralized IR functions that can adapt and scale in continuously growing and increasingly interconnected information spaces.

Zhao Li received the PhD (hons.) degree from the Computer Science Department, University of Vermont, Burlington, Vermont. He is currently a senior staff scientist with the Alibaba Group, specializing in ecommerce ranking and recommendation systems. He has published several articles in prestigious conferences and journals, including NIPS, AAAI, IJCAI, SIGIR and the *IEEE Transactions on Knowledge and Data Engineering*.