Multi-Network Graph Contrastive Learning for Cancer Driver Gene Identification

Wei Peng, *Member, IEEE*, Zhengnan Zhou, Wei Dai, Ning Yu, and Jianxin Wang, *Senior Member, IEEE*

***Abstract*—Identifying driver genes contributing to the occurrence and development of cancers plays a critical role in cancer research and treatment. Some recent computational approaches identify cancer-driver genes based on gene networks, assuming that cancer-driver genes perform essential functions in gene networks. Due to the noise in gene function networks, many works focus on integrating gene networks derived from multi-omics datasets to improve the accuracy of cancer driver gene detection. However, most of them ignore the information interactions between these multi-omics datasets. In this work, we propose MNGCL, a Multi-Network Graph Contrastive Learning method to identify cancer driver genes. It first constructs three gene networks as different views based on protein interactions, gene semantic similarities, and gene co-occurrence in signaling pathways. Then, we perform data augmentation of these gene networks and input them into a graph contrastive learning (GCL) encoder with shared parameters to learn consistent gene feature representation in different networks from a holistic perspective. After that, the gene features from the GCL encoder are passed through three different graph convolutional networks to generate the unique gene feature representations in the three networks. Finally, we used a logistic regression model to fuse the gene feature representations generated in each network to predict cancer driver genes. The experimental results show that MNGCL improves the area under the ROC curve (AUC) and the area under the precision-recall curve (AUPRC) to a greater extent than the existing methods in identifying driver genes for both pan-cancer and single-type cancers. Furthermore, the ablation studies show that our model capturing dependencies and interactions between gene networks provided a more comprehensive perspective on the molecular mechanisms underlying cancer and improved the accuracy of cancer driver identification.**

Manuscript received 5 September 2023; revised 31 January 2024; accepted 2 March 2024. Date of publication 8 March 2024; date of current version 12 June 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 61972185, in part by the Natural Science Foundation of Yunnan Province of China under Grant 2019FA024, and in part by Yunnan Ten Thousand Talents Plan young. Recommended for acceptance by Dr. Y. Liu. *(Corresponding authors: Wei Peng; Jianxin Wang.)*

Wei Peng, Zhengnan Zhou, and Wei Dai are with the Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650093, China, and also with the Computer Technology Application Key Lab of Yunnan Province, Kunming University of Science and Technology, Kunming 650093, China (e-mail: weipeng1980@gmail.com; alen12nan@gmail.com; daiwei@kust.edu.cn).

Ning Yu is with the Department of Computing Sciences, College at Brockport, State University of New York, Brockport, NY 14422 USA (e-mail: nyu@brockport.edu).

Jianxin Wang is with the School of Computer Science and Engineering, Central South University, Changsha 410083, China, and also with the Hunan Provincial Key Lab on Bioinformatics, Central South University, Changsha 410083, China (e-mail: jxwang@mail.csu.edu.cn).

The source code can be obtained from https://github.com/weiba/MNGCL.

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSE.2024.3373652, provided by the authors.

Digital Object Identifier 10.1109/TNSE.2024.3373652

***Index Terms*—Cancer driver genes, multi-view gene network, graph contrastive learning, network integration.**

I. Introduction

CANCER is a kind of genomic disease. Mutations in certain genes can contribute to unregulated cell proliferation, leading to cancer. Those genes that promote cancer progression are called cancer driver genes [1], [2]. Proper identification of cancer driver genes plays a critical role in understanding the molecular mechanisms of cancer development, developing personalized treatment and designing anti-cancer drugs. In recent years, large-scale cancer genomics projects, such as The Cancer Genome Atlas (TCGA) [3], the International Cancer Genome Consortium (ICGC) [4], and the Catalogue of Somatic Mutations in Cancer (COSMIC) [5], published genomic, transcriptomic, epigenomic, and proteomic data from tens of thousands of cancer patients. High-throughput experimental techniques have also generated a large amount of gene function and protein interaction data. These data have been widely used by researchers to design computational methods to identify cancer driver genes [6], [7], [8], [9], [10].

In this paper, we design a Multi-Network Graph Contrastive Learning method to identify cancer driver genes called MNGCL by integrating multiple gene networks. It first constructs three gene networks as different views of genes based on protein-protein interactions, gene semantic similarities, and gene co-occurrence in signaling pathways. Then, MNGCL performs data augmentation on these gene networks and inputs them into a graph contrastive learning (GCL) encoder with shared parameters to learn the feature representations of genes in each network. After that, the gene features from the GCL encoder are passed through three different graph convolutional networks to generate the gene feature representations in the three networks. Finally, we use a logistic regression model to fuse the gene feature representations generated in each network to predict cancer driver genes. The experimental results show that MNGCL improves the area under the ROC curve (AUC) and the area under the precision-recall curve (AUPRC) to a greater extent than the existing methods in identifying driver genes for both pan-cancer and single-type cancers.

The main contributions of this paper are summarized as follows:

- To the best of our knowledge, MNGCL is the first attempt to apply a multi-graph contrastive learning approach to represent gene features and predict cancer driver genes. It

regards the three different gene networks as three different views of genes. From the node level, each gene learns new feature representations through Chebyshev graphs to converge neighbor features in corresponding gene network. From a view level, the method integrates information from multiple networks while constraining and interacting between different networks to preserve global-level and more reliable gene features to identify cancer driver genes.

- We optimize the contrastive loss by adjusting positive sampling strategies during contrastive learning. Since genes play different roles in different networks, we input the gene features learned in the multi-graph contrastive learning into a network-specific Chebyshev graph convolutional encoder to learn the unique gene feature representations in the respective networks.
- We apply MNGCL to predict cancer driver genes for pan-cancer, single types of cancer, and independent datasets. Numerous experimental results show that our method can learn the global feature representation of genes in multiple networks due to multi-graph contrastive learning and thus can predict cancer driver genes more accurately than existing baseline methods.

II. Related Work

A. Cancer Driver Gene Identification

Computational methods for identifying cancer driver genes fall into two main categories. One category is the mutation frequency-based methods. The other category is biological network-based methods. The earliest methods tend to compare the mutation frequency of cancer driver genes with a predefined background mutation rate based on the principle that cancer driver genes have a higher mutation frequency than passenger genes. We called these methods, such as OncodriveCLUST [11] and ActiveDriver [12], as frequency-based methods. However, most of these methods do not correctly estimate the background mutation rate, thus ignoring cancer driver genes with low mutation frequencies. Biological networks, as abstract representations of biological systems, are rich in semantic knowledge. Therefore, several methods based on protein-protein interaction (PPI) networks have emerged. For example, HotNet2 [13] performs heat diffusion modeling on PPI networks to detect cancer driver gene modules with high mutation characteristics. MUFFINN [14] ranks genes by considering the effect of mutation frequencies of neighbors in the PPI network and recommends the top-ranked ones as cancer driver genes. However, there is a large amount of false positive data in the network. To address this issue, some literature [15], [16], [17], [18], [19] uses multi-omics data to weigh the PPI network, removing noisy edges under the constraints of inter-molecular co-expression, co-function, co-subcellularity, and co-tissue. However, these methods consider more associations between genes and do not take full advantage of gene characteristics to predict cancer-driver genes. DeepDriver [20] extracts 12 features from gene DNA sequences that are closely related to cancer development and then splices the features of each gene with the features of its K nearest neighboring nodes in the co-expression network to form a feature matrix of each gene, and then input the gene feature matrix into a convolutional neural network to predict cancer driver genes. RLAG [21] constructs a gene attribute network according to subcellular information and a gene structure network according to gene functional information and then executes node2vec on the two gene networks to learn the gene features and predict the cancer driver genes.

In recent years, graph neural network (GNN) models have been widely used in the field of bioinformatics because they can simultaneously consider the node attributes and their network structure, when learning node features [22], [23], [24], [25], [26], [27]. EMOGI [22], MTGCN [23] and HGDC [24] are based on Graph Convolutional Network (GCN) [28] frameworks that combine genomics, epigenomics, and transcriptomics data as gene features with PPI networks to learn low-dimensional representations of network node features, and achieve good results in cancer driver gene prediction tasks. However, all these approaches use only PPI networks. MODIG [25] introduces multiple types of gene relationships (including gene functional similarity, gene co-expression patterns, etc.) to construct different gene-gene networks, and then proposes a graph attention neural network model to integrate gene features in multiple networks to predict cancer driver genes. Peng et al. [26] consider that tumorigenesis usually involves complex interactions between genes and other molecules, thus construct gene-outlying gene networks and gene-miRNA networks, and propose a new method called the MRNGCN to identify cancer driver genes based on multi-gene relational networks and heterogeneous graph convolution models. Although multi-gene network data can complement each other to characterize genes, these methods overlooked the constraints and information interactions between different gene networks.

B. Graph Contrastive Learning

Contrastive learning is a highly promising unsupervised learning method that has achieved significant success in both the computer vision (CV) and natural language processing (NLP) domains. Recently, contrastive learning has attracted considerable attention in many network representation learning tasks [29], [30], [31]. The main idea is to generate different representations of the same sample and use contrastive learning loss to maximize their agreement while minimizing the similarity between other negative samples. In a typical Graph Contrastive Learning (GCL) framework, the first step involves obtaining a novel view through various data augmentation techniques. For instance, GRACE [32] achieves data augmentation by uniformly deleting edges in the network and masking feature attributes. You et al. [31] proposed an approach that leverages node dropping to obtain an augmented view. Meanwhile, GRADATE [33] and MSSGCL [34] argue that using subgraph sampling as a data augmentation method can better preserve the semantic knowledge in the original view.

III. Materials

The multi-omics data we used were obtained from TCGA, which contains gene mutations, DNA methylation, and gene

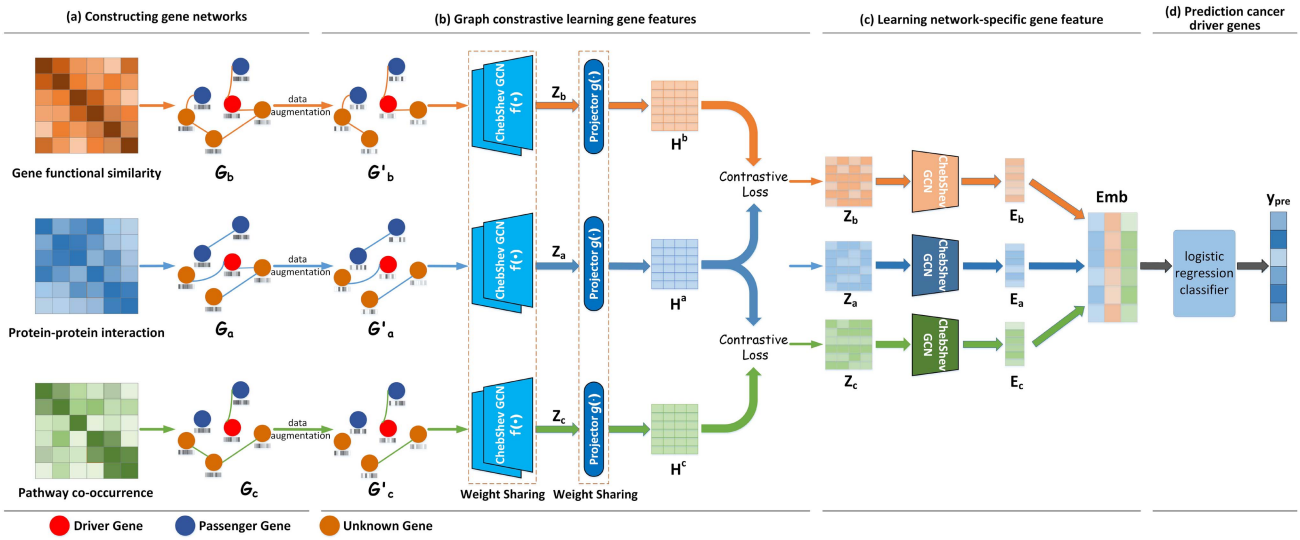


Fig. 1. Framework of MNGCL.(a) It constructs three gene networks as different views of genes based on protein-protein interactions, gene semantic similarities, and gene co-occurrence in signaling pathways.(b) performs data augmentation on multiple networks, and then trains an encoder gene features with shared parameters by contrastive learning. (c) Obtains network-specific gene feature representations through different encoders and fuses them to generate Embeddings. (d) uses a logistic regression model to predict cancer driver genes.

expression. The gene expression data was from Wang et al. [35], and was obtained by ComBat [36] for further normalization and batch correction. As with EMOGI [22] and MTGCN [23], we only focused on cancer types for which gene mutations, gene expression data and DNA methylation information are available in both tumor and normal tissues. Therefore, this work covers 16 cancer types and more than 8,000 samples. The PPI data were obtained from CPDB [37] and STRING [38]. After removing interactions with scores < 0.5 in CPDB, we obtained a PPI network of CPDB, including 13,627 nodes and 504,378 edges. Similarly, after removing the interactions with a score of < 0.85 in STRING, we obtained a PPI network of STRING consisting of 13,179 nodes and 673,099 edges.

We downloaded the gene semantic similarity network and the pathway co-occurrence network from the supplementary file of MODIG [25]. We performed some processing on them and only kept the same gene nodes as in the PPI network, removed interactions with functional similarity < 0.8 in the gene semantic similarity network, removed interactions with co-occurrence relationship < 0.6 in the pathway co-occurrence network. Therefore, if we took the CPDB PPI network, we got 13,627 nodes with 881,038 edges in the gene semantic similarity network and 13,627 nodes with 261,312 edges in the gene pathway co-occurrence network. If we selected the STRING PPI network, we got 13,179 nodes with 602,600 edges in the gene semantic similarity network and 13,179 nodes with 426,336 edges in the gene pathway co-occurrence network. Note that there are some isolated nodes in the two networks. We downloaded pan-cancer-positive samples from the attachment of MTGCN, which contains the high-confidence driver genes from the Network of Cancer Genes (NCG 6.0) [39], COSMIC Cancer Gene Census (CGC v91) [5], and DigSEE [40]. We started from all genes and recursively removed genes in NCG, COSMIC, Online Mendelian Inheritance in Man (OMIM) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway to obtain a negative sample list. Thus, our dataset included 796 positive and 2,187 negative samples.

IV. Method

A. Overview

As shown in Fig. 1, MNGCL is a multi-network contrastive learning framework for cancer driver gene identification. First, MNGCL constructs multiple gene-gene relationship networks, and these different networks describe the relationships among genes from different views. Then, it performs contrastive learning on different relationship networks to learn consistent gene feature representation in different networks from a holistic perspective. Thirdly considering that genes play different roles in different networks, we input the gene features learned in the previous step into each network-specific Chebyshev graph convolutional encoder to learn the unique feature representations of genes in the respective networks. Finally, we pass the unique features of genes learned from the three networks through a logistic regression classifier for the downstream cancer-driven gene identification task. For easy reference,Table I summarizes the key notations and their definitions.

B. Construction of Gene Networks

We constructed three gene-gene association networks through protein-protein interactions, gene functional similarity, and gene co-occurrence in signaling pathways. The values of these association profiles are in the range of [0,1].

1) Protein-Protein Interaction Network: Protein-protein interaction (PPI) networks provide a comprehensive description of protein interactions by participating in various aspects of life processes such as biological signaling, regulation of gene expression, energy and material metabolism, and cell cycle regulation. Through the PPI network, we constructed the gene-gene

TABLE I
KEY NOTATIONS AND DEFINITIONS

Notation	Definition
G_a	The protein-protein interaction network
G_b	The gene functional similarity network
G_c	The pathway co-occurrence association network
X	The gene feature matrix
p	The set of subscripts of G_a, G_b, G_c, i.e. $\{a, b, c\}$
A_p	The adjacency matrix of G_p
G'_p	The data enhanced networks of G_p
Z_p	The feature representation of G'_p after encoder
H_p	The projection matrix of G_p
h_i^p	The feature vector of gene i in the projection matrix H_p
ω	The learnable parameters
N	The number of gene nodes in the network
N_i^p	The set of neighboring nodes of node i in G_p
τ	The temperature hyperparameter
y_i	The true label of the gene i
E_p	The final gene representation in G_p
U	The number of encoders to be constrained during training
λ	The weight between the two losses
$\hat{y}$	The prediction score

relationship network and its adjacent matrix. In the network, we used the same method as MTGCN [23] for the initial features of genes. The initial features of genes are composed of biological features and network topological features. The biological features were obtained by calculating the gene mutation rate, differential DNA methylation rate, and differential gene expression rate for each cancer type. Since we focused on only 16 cancer types, each gene has a 48-dimensional biometric signature consisting of 16 mutation rates, 16 methylation values, and 16 differential expression rates, obtained after min-max normalization. By using the node2Vec algorithm on, we obtained the 16-dimensional topological features of the genes. We concatenated the biometric and topological features to obtain the 64-dimensional initial feature matrix X of the gene.

2) Gene Functional Similarity Network: The gene functional similarity network was obtained by measuring the functional similarity of each gene pair based on the semantic similarity between GO terms of annotated genes. Specifically, using the R package GOSemSim [41] in the mgeneSim function, the functional similarity of each pair of genes can be calculated and the gene functional similarity network can be constructed by keeping the interactions in the network with functional similarity > 0.8. Similarly, the initial features of gene nodes in this network are 64-dimensional features X.

3) Pathway Co-Occurrence Association Network: Pathways may involve multiple genes, proteins, and small molecules that together participate in particular biological processes or functions.The pathway co-occurrence network describes the associations and potential functional relationships of genes in cancer-related signaling pathways based on their co-occurrence patterns within biological pathways. We use the human cancer pathways in the KEGG database to construct a 337-dimensional binary vector for each gene, 1 if the gene appears in the corresponding pathway and 0 otherwise. Then, the gene co-occurrence relationship between two genes was calculated by cosine similarity between two gene vectors. In this way, we constructed the gene-gene network obtained from the gene pathway co-occurrence relationships by keeping the interactions with co-occurrence relationship > 0.6 in the network. Similarly, the initial feature matrix of gene nodes in this network is X.

C. Chebyshev GCN Encoder

Throughout our approach, the Chebyshev GCN model [42] are chosen as the feature encoders. Compared to the GCN, the Chebyshev GCN has higher constraints on aggregating information from higher-order neighbors and also reduces computational complexity. Given the adjacency matrix A of any gene network and the gene feature matrix X, a single Chebyshev convolutional layer is defined as follows:

$$\mathrm{H} = \sigma\left(\sum_{\mathrm{k}=1}^{\mathrm{K}} \mathrm{Z}^{(\mathrm{k})} \cdot \theta^{(\mathrm{k})}\right), \tag{1}$$

where H is the feature to be learned by the Chebyshev GCN layer, σ is the activation function, such as ReLU and Sigmoid, θ is the matrix of learnable parameters in the neural network, and $Z^{(k)}$ is computed by recursive methods:

$$\begin{aligned} Z^{(1)} &= X \\ Z^{(2)} &= \hat{L} \cdot X \\ Z^{(k)} &= 2 \cdot \hat{L} \cdot Z^{(k-1)} - Z^{(k-2)}, \end{aligned} \tag{2}$$

where K represents the size of the Chebyshev GCN. Here, we set $K = 2$, $\hat{L} = \frac{2L}{\lambda_{\max}}$ represents the scaled and normalized Laplace operator. We set $D_{ii} = \Sigma_j A_{ij}$, $L = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$. Here, I is the unit matrix. $\lambda_{\max}$ is the maximum eigenvalues of L.

D. Contrastive Learning

As described in Section 'Construction of gene networks', we construct three gene-gene networks, i.e.,G_a, G_b and G_c with corresponding adjacency matrices A_a, A_b and A_c, respectively. Their gene initial feature matrices all are X. Then, we perform data augmentation on each network, and put the data-augmented networks into an encoder with shared parameters for contrastive learning.

1) Data Enhancement: We first randomly crop the edges in the networks G_a, G_b and G_c with different cropping probabilities in each network. Then we randomly mask the values of some attributes in the feature matrix X. The feature masking probabilities are also different in each network. This way we generate the data enhanced networks G'_a, G'_b and G'_c. The adjacency matrix and gene features of the augmented networks are the initial features of our next input encoder.

2) Contrastive Learning of Coding Gene Features: Here, our goal is to learn consistent gene feature representations under different views. At the node level, each gene learns a new feature representation by aggregating neighboring node features in the corresponding relational network through Chebyshev graph convolution network with shared weights. At the view level, we make full use of the interaction information between different views by comparing three different gene networks to learn consistent gene feature representations, which can effectively reduce the impact of noise in each network. Specifically, we

learn the feature representations Z_a, Z_b and Z_c for each of the three networks G'_a, G'_b and G'_c after inputting them into a Chebyshev GCN feature encoder with shared weights. To mitigate the information loss due to contrastive loss, we map the individual gene feature representations onto a common space via a two-layer nonlinear projection head with shared parameters:

$$H^p = \omega^{(2)}(ELU(\omega^{(1)}Z_p)), \tag{3}$$

where $p \in \{a, b, c\}$, $\omega^{(1)}$ and $\omega^{(2)}$ represent the learnable parameters in the first and second layers of the fully connected neural network, respectively. a, b, c represent three different gene relationship networks. The ELU is Elu activation function. Next, we perform contrastive learning between the projected gene representations. In the contrastive loss process, h_i^p represents the feature vector of gene i in the projection matrix H^p. We first compute the cosine similarity of the projected feature representations of the gene i in a particular network a and in other networks b, as follows:

$$sim(h_i^a, h_i^b) = \frac{(h_i^a)(h_i^b)^{\mathrm{T}}}{\|h_i^a\| \left\|h_i^b\right\|}. \tag{4}$$

When calculating the contrastive loss, positive and negative sample pairs need to be defined. The key idea behind contrastive learning is to learn data representations that maximize similarities of positive pairs while minimizing those of negative pairs. In traditional contrastive learning, only the same samples with different data are generally considered as positive sample pairs, and the rests are negative sample pairs. However, this may ignore the information from neighborhood nodes in the network. Considering that genes in the network usually perform similar functions to their local context, we propose a new strategy where all genes with interactions in the network are mutually considered as positive pairs. Otherwise are negative pairs. For example, when comparing G_a with G_b, we have the following contrastive loss:

$$l_i^{ab} = -\log \frac{\sum_{j \in N_i^a} \exp(sim(h_i^a, h_i^b)/\tau + sim(h_i^a, h_j^a)/\tau)}{\sum_{k=1}^{N} \exp(sim(h_i^a, h_k^b)/\tau)}, \tag{5}$$

$$l_i^{ba} = -\log \frac{\sum_{j \in N_i^b} \exp(sim(h_i^b, h_i^a)/\tau + sim(h_i^b, h_j^b)/\tau)}{\sum_{k=1}^{N} \exp(sim(h_i^b, h_k^a)/\tau)}, \tag{6}$$

where N_i^a and N_i^b are the sets of neighboring nodes of node i in G_a and G_b, respectively. The τ is the temperature hyperparameter, which τ is 0.3 in our method. Similarly, we can get the contrastive loss, l_i^{ac} and l_i^{ca}, when comparing G_a with G_c. We only let H^b and H^c do contrastive learning with H^a respectively. Because a large number of different kinds of protein mutual information are contained in PPI, and previous studies have shown that the structural properties of genes in PPI networks are effective for identifying cancer driver genes [21]. Therefore, in the contrastive learning phase, our overall objective function is:

$$l_{con} = \frac{1}{N} \sum_{i=1}^{N} (l_i^{ab} + l_i^{ba} + l_i^{ac} + l_i^{ca}), \tag{7}$$

E. Unique Feature Representation of Genes in the Three Networks

We learn the mutual information between networks by performing contrastive learning across multiple networks, taking advantage of the differences between them to draw closer to the feature representation of the same node in different networks. However, considering the unique role of genes in each network, we utilize three different Chebyshev encoders to extract gene features in the three networks. Specifically, we use three single-layer Chebyshev GCN encoders with unshared parameters to input Z_a, Z_a, Z_c and their corresponding enhanced network adjacency matrices into the encoders. Here, we use a single-layer convolutional layer to reduce the features to one dimension and obtain the feature embedding representations of the genes in each network E_a, E_b, E_c, respectively. In this process, we use a binary cross-entropy loss to constrain each encoder, i.e:

$$l_{pre\_p} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log \sigma(E_p) + (1 - y_i) \log(1 - \sigma(E_p))], \tag{8}$$

where $p \in \{a, b, c\}$, σ is the sigmoid activation function, y_i is the true label (0 or 1) of the gene i, and N is the number of gene nodes in the network.

F. Model Training and Prediction Task

1) Model Training: In our model, we jointly optimize the two losses of (7) and (8), which can be expressed as follows:

$$l_{total} = (1 - U\lambda) l_{con} + \lambda \sum_{p=1}^{U} l_{pre\_p}, \tag{9}$$

where U is the number of encoders to be constrained during training. Since we input three networks in this work, we have $U = 3$. The $\lambda \in [0, \frac{1}{U}]$ is a manually set hyper-parameters to adjust the weight between the two losses.

2) Classification Prediction of Cancer Driver Genes: As mentioned in the previous section, we learn a final feature representation for each gene by re-inputting the features learned after contrastive learning for each network into each network-specific Chebyshev graph convolutional encoder, namely E_a, E_b, E_c. We put these features, rich in information about each biological network, into a logistic regression model to implement the classification prediction of cancer driver genes. In a general form, the logistic regression model is defined as follows:

$$x = \omega_1 E_a + \omega_2 E_b + \omega_3 E_c + b, \tag{10}$$

$$\hat{y} = \frac{1}{1 + e^{-x}}, \tag{11}$$

where b is the number of biases that can be learned in the model. ω_1, ω_2 and ω_3 are the learnable weight parameters in the logistic regression model, which represent the contribution of the

Algorithm 1: MNGCL Method.

Input: Feature matrix X; adjacency matrices A_a, A_b, A_c; feature masking rate m_a, m_b, m_c; edge dropout rate d_a, d_b, d_c; loss weight λ; number of epochs e.
Output: Prediction score $\hat{y}$.
while trained epochs $< e$ **do**
  $X'_a = Featuremask(X, m_a)$;
  $X'_b = Featuremask(X, m_b)$;
  $X'_c = Featuremask(X, m_c)$
  $A'_a = Edgedropout(A_a, d_a)$;
  $A'_b = Edgedropout(A_b, d_b)$;
  $A'_c = Edgedropout(A_c, d_c)$
  $Z_a = f(X'_a, A'_a)$ // Chebyshev GCN encoder $f(\cdot)$
  $Z_b = f(X'_b, A'_b)$
  $Z_c = f(X'_c, A'_c)$
  $H^a = \omega^{(2)}(ELU(\omega^{(1)} Z_a))$
  $H^b = \omega^{(2)}(ELU(\omega^{(1)} Z_b))$
  $H^c = \omega^{(2)}(ELU(\omega^{(1)} Z_c))$
  Calculating the Contrastive Loss by (4)–(7).
  $E_a = Chebshev(Z_a, A'_a)$
  $E_b = Chebshev(Z_b, A'_b)$
  $E_c = Chebshev(Z_c, A'_c)$
  Calculating the Binary Cross Entropy Loss of each network by (8)
  Calculating the l_{total} by (9)
  Update the weights of MNGCL by l_{total}
  $Emb = Concat(E_a, E_b, E_c)$
  $\hat{y} = LogReg(Emb)$
end while
return $\hat{y}$

properties of genes features in each network to the identification of cancer driver genes, respectively. In this way, we output for each gene a prediction score of whether or not it is a driver gene. Algorithm 1 lists the pseudocode for running our model.

V. Experiment and Result Analysis

A. Baseline

To evaluate the performance of our model, we chose nine baseline methods, including (GCN [28], Chebnet [42], EMOGI [22], MTGCN [23], BIONIC [43], MODIG [25], MRNGCN [26], HGDC [24] and MRNGCN-MG), to compare with our mode. Among them, GCN is a typical graph neural network approach that learns new features by aggregating features from its direct neighbors and itself. Chebnet is a graph convolutional neural network model using the Chebyshev GCN encoder. EMOGI and MTGCN are the latest deep learning methods based on graph convolution, which use multi-omics data as gene features and combine PPI networks to learn gene features to predict cancer driver genes. HGDC is a novel method for identifying cancer driver genes based on the PPI network and graph diffusion convolution. They employ graph diffusion techniques to create an auxiliary network on the PPI network for cancer driver gene identification. It combines features learned from the PPI and the auxiliary network but does not involve contrastive learning. BIONIC is a GAT-based network integration algorithm that integrates multiple numbers and sizes of networks to learn gene features. Here, we use our three networks as input of BIONIC to predict cancer driver genes. MODIG and MRNGCN are both the latest multi-relationship network integration methods. MODIG is a graph attention network (GAT)-based framework that combines five gene relationships, protein-protein interactions, gene sequence similarity, gene pathway co-occurrence, gene co-expression patterns, and gene functional similarity to identify cancer driver genes. MRNGCN is a method that uses heterogeneous graph convolution and self-attentiveness to fuse genes in the PPI network, gene-outlying gene network, and gene-miRNA network to identify cancer driver genes. MRNGCN-MG is a variation of MRNGCN in which we changed the input to our three networks and replaced its heterogeneous graph convolution with the same Chebyshev graph convolution as MNGCL in our model. We used the same PPI network and initial gene feature matrix in each method for a fair comparison. Considering most previous works evaluate their works under ten five-fold cross-validation [23], [25], [26], we fixed the hyperparameter setting of all models and compared our model with baselines to predict pan-cancer driver genes and cancer type-specific driver genes using the average results of ten five-fold cross-validations. For five-fold cross-validations, we randomly divided the labeled data into equal five parts with four parts as training set (80%) and the rest parts as test set (20%). In these two sets, the proportion of known cancer driving genes and non cancer driving genes is the same as that of the original dataset.

B. Parameter Settings

Our algorithm is implemented in Python 3.7, PyTorch 1.9.1, and Pytorch geometric 2.0.4 environment, and the chosen optimizer is Adam. We ran a grid search within a reasonable parameter range to find the optimal hyperparameter combinations of our model. We first trained a MNGCL model on the CPDB PPI network for each possible hyperparameter combination, corresponding to a total of 24 combinations (Leraning rate ranging in [0.001,0.002], EPOCH ranging in [301,501,1001], λ ranging in [0.1,0.15,0.2,0.25]). We then used ten five-fold cross-validation on the training set to obtain a robust result independent of the random initialization of the network weights. We chose the combination with the highest average AUPRC (the 9th combination in Fig. 2) with a learning rate of 0.001, an epoch of 1000 and a loss weight $\lambda = 0.1$.

Therefore, the optimal combination of hyperparameters is as follows: learning rate of 0.001, epoch of 1,000, the loss weight $\lambda = 0.1$. In the data enhancement phase, the edge dropout rate of PPI network, gene semantic similarity network and pathway co-occurrence association network are 0.2, 0.1, 0.1 for CPDB dataset and 0.5, 0.3, 0.2 for STRING dataset, and feature masking rate are 0.5, 0.3, 0.4 of three networks on the CPDB and STRING datasets. In the contrastive learning encoding gene features phase, we used two layers of Chebyshev GCN with filter sizes of 300, 100, respectively. In the cancer type-specific experiments, considering that the input features were reduced

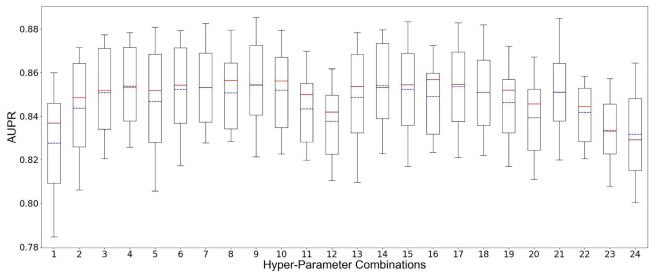


Fig. 2. Grid search for best-performing combinations of hyper-parameters.

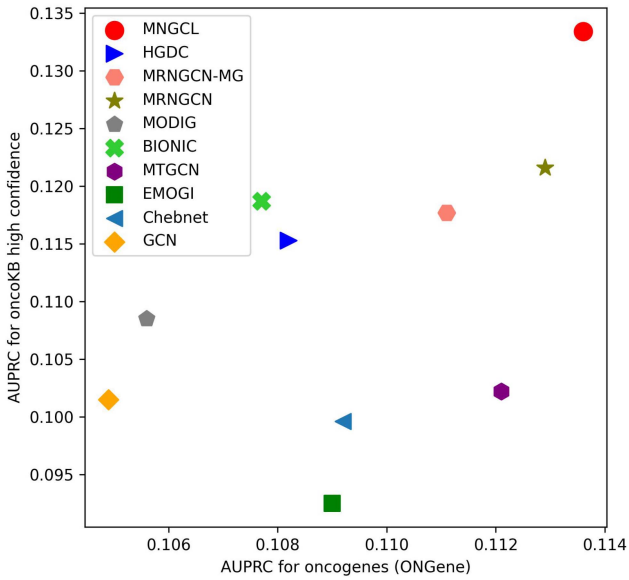


Fig. 3. Performance comparisons of the different methods on two independent cancer driver gene sets.

to 19 dimensions, we reduced the feature masking rate of gene nodes in each network to 0.3, 0.1, and 0.2, respectively, with an epoch of 1,200. The size of the Chebyshev layers for the contrastive learning encoding gene features stage was set to 150, 50, respectively, and the rest parameters were kept the same as in the pan-cancer experiments. All parameters of the baseline method were set under the recommendations in the paper or adjusted appropriately to achieve the best performance.

C. Performance on Pan-Cancer Driver Gene Prediction

We applied our method and baselines to predict pan-cancer driver genes and used the average of AUC and AUPRC under ten five-fold cross-validation as evaluation metrics. As shown in Table II, our method achieved the best performance on the CPDB and STRING datasets,with AUC and AUPRC values reached 93.17%, 85.75% and 93.31%, 84.92%, respectively, demonstrating the validity of MNGCL. The results show that MODIG and MRNGCN improve performance over previous methods after using multi-relational networks. Among them, MODIG uses many networks to obtain more information. Meanwhile, MRNGCN-MG used the same input networks and encoders as ours. However, their performance is much lower than that of MNGCL. Ours outperforms baseline attributes to the effectiveness of our model learning mutual information between different networks.

TABLE II
PERFORMANCE COMPARISON ON PAN-CANCER DRIVER GENE PREDICTION

	CPDB		STRING	
Methods	AUC	AUPRC	AUC	AUPRC
GCN	0.8855	0.7709	0.8718	0.7149
Chebnet	0.9041	0.8230	0.9043	0.8005
EMOGI	0.9044	0.8169	0.9097	0.8078
MTGCN	0.9116	0.8332	0.9168	0.8189
BIONIC	0.9169	0.8250	0.9149	0.8034
MODIG	0.9182	0.8360	0.9251	0.7933
MRNGCN	0.9192	0.8446	–	–
MRNGCN-MG	0.9151	0.8296	0.9030	0.7933
HGDC	0.9123	0.8309	0.8954	0.7642
MNGCL	**0.9317**	**0.8575**	**0.9331**	**0.8492**

D. Performance on Cancer Type-Specific Driver Gene Prediction

We also investigated the effectiveness of MNGCL in detecting driver genes for single cancer types on CPDB dataset, including breast invasive carcinoma (BRCA), lung adenocarcinoma (LUAD), bladder urothelial carcinoma (BLCA), liver hepatocellular carcinoma(LIHC), cervical squamous cell carcinoma and endocervical adenocarcinoma(CESC), colon adenocarcinoma(COAD), esophageal carcinoma(ESCA), head and neck squamous cell carcinoma(HNSC), kidney renal clear cell carcinoma(KIRC), kidney renal papillary cell carcinoma(KIRP), lung squamous cell carcinoma(LUSC), prostate adenocarcinoma(PRAD), stomach adenocarcinoma(STAD), thyroid carcinoma(THCA) and uterine corpus endometrial carcinoma(UCEC). Positive samples for a single cancer type were obtained from NCG6.0 labeled with that cancer type. The negative sample consisted of 2,187 genes, the same as the pan-cancer data. For a single cancer type, we reorganized the initial attributes of the nodes of these three networks. The initial gene attributes for each cancer type were 19-dimensional and included three biological features of that cancer type (gene mutation rate, differential DNA methylation rate, and differential gene expression rate) and 16-dimensional structural features learned from the gene networks. In terms of networks, we used the same three networks for pan-cancer.

As seen in Table III, in 15 cancer types, the AUPRC values of MNGCL were much higher than other methods except KIRP. For example, it outperformed the second-best method MRNGCN by 6.42% and 7.52% for the larger sample sizes BRCA (201 positive samples) and HNSC (141 positive samples), respectively. In UCEC (68 positive samples), which has a medium sample size, it similarly outperforms the suboptimal method by 5.65%. In CESC (16 positive samples), which has a very small sample size, it outperforms the second-best method MTGCN by 3.30%. All these experimental results indicate that MNGCL still performs

TABLE III
PERFORMANCE OF CANCER TYPE-SPECIFIC DRIVER GENE PREDICTION

Cancer type	LUAD	BRCA	BLCA	LIHC	CESC	COAD	ESCA	HNSC	KIRC	KIRP	LUSC	PRAD	STAD	THCA	UCEC
Model/AUC															
GCN	0.8042	0.8813	0.8712	0.8098	0.8201	0.7630	0.8129	0.7831	0.7669	0.7012	0.7187	0.8304	0.8176	0.7725	0.8416
Chebnet	0.8841	0.9020	0.9375	0.8801	0.9713	0.8094	0.8809	0.8224	0.8626	0.9404	0.8971	0.9007	0.8935	0.8212	0.8756
EMOGI	0.8709	0.8989	0.9359	0.8753	0.9740	0.8060	0.8673	0.8131	0.8992	0.9351	0.8411	0.9094	0.8883	0.7975	0.8713
MTGCN	0.9019	0.9061	0.9495	0.8937	**0.9805**	0.8295	0.9020	0.8387	0.9078	0.9645	0.8984	0.9170	0.9265	0.8263	0.9066
BIONIC	0.9048	0.9173	0.9336	0.8940	0.9612	0.8604	0.9038	0.8577	0.9218	0.9388	0.9374	0.9617	0.9165	0.8348	0.9075
MODIG	0.9192	0.9106	0.9504	0.8922	0.9524	0.8486	0.9167	0.8517	0.8917	0.9612	**0.9570**	0.9491	0.9341	0.8290	0.9014
MRNGCN	**0.9427**	0.9120	0.9544	0.9109	0.9627	0.8502	0.9187	0.8695	0.9032	**0.9688**	0.9273	0.9282	0.9222	0.8347	0.9061
MRNGCN-MG	0.8905	0.9006	0.9426	0.8730	0.9671	0.7968	0.8952	0.8128	0.8713	0.9408	0.9104	0.9292	0.9066	0.8143	0.8671
HGDC	0.8908	0.9041	0.9406	0.8919	0.9757	0.8136	0.8768	0.8059	0.8758	0.9300	0.8869	0.9009	0.9098	0.8330	0.8770
MNGCL	0.9175	**0.9259**	**0.9641**	**0.9166**	0.9706	**0.8738**	**0.9350**	**0.8847**	**0.9227**	0.9546	0.9146	**0.9505**	**0.9428**	**0.8355**	**0.9312**
Model/AUPRC															
GCN	0.4187	0.5866	0.3191	0.3017	0.0980	0.2680	0.2959	0.3290	0.0861	0.0235	0.0401	0.3010	0.2502	0.1519	0.2550
Chebnet	0.5771	0.6507	0.5394	0.4215	0.5261	0.3490	0.4288	0.4026	0.2750	0.2420	0.2401	0.4910	0.4730	0.2860	0.4021
EMOGI	0.5591	0.6482	0.5485	0.3845	0.5547	0.3459	0.4174	0.3999	0.3438	0.2449	0.2222	0.5233	0.4604	0.2587	0.4002
MTGCN	0.6279	0.6583	0.6568	0.4645	0.5972	0.3816	0.4721	0.4396	0.3498	0.4358	0.3099	0.6143	0.5931	0.2907	0.4954
BIONIC	0.6129	0.6911	0.6145	0.5038	0.5157	0.4437	0.4954	0.5010	0.3123	0.3372	0.3050	0.7045	0.5468	0.2850	0.4691
MODIG	0.6916	0.6809	0.6595	0.4675	0.3950	0.4336	0.4898	0.4733	0.3542	0.4702	0.3865	0.6285	0.6251	0.2699	0.4379
MRNGCN	0.6287	0.6920	0.7248	0.5468	0.5339	0.4953	0.5366	0.5442	0.4097	**0.5967**	0.3478	0.6688	0.6458	0.3080	0.5697
MRNGCN-MG	0.6404	0.6507	0.6337	0.4239	0.4498	0.3322	0.4357	0.4144	0.3303	0.3025	0.3426	0.5870	0.5681	0.2612	0.3636
HGDC	0.6062	0.6477	0.6049	0.4364	0.5152	0.3445	0.4129	0.3892	0.2909	0.1721	0.2331	0.5003	0.5190	0.2943	0.3732
MNGCL	**0.7093**	**0.7562**	**0.7732**	**0.5605**	**0.6302**	**0.5512**	**0.5844**	**0.6194**	**0.4450**	0.4444	**0.4291**	**0.7466**	**0.6836**	**0.3102**	**0.6259**

The best results are highlighted in bold.

best in individual cancer data with imbalanced positive and negative samples.

E. Ablation Experiments

MNGCL is a method based on a graph contrastive learning framework that uses multiple different gene views for contrastive learning, aiming to learn the mutual information of genes under different views. Our model used three gene relationship networks obtained from CPDB protein-protein relationships (PPI), gene semantic functional similarity, and gene pathway co-occurrence. We performed ablation experiments on pan-cancer dataset to verify that the information brought by the different views was valid. Where "PPI + PPI" indicates just using the PPI network, two different gene views were obtained using different enhancement strategies (edge dropout rate of 0.2, 0.1, and feature masking rate of 0.5, 0.3, respectively). Similarly, "GO + GO" uses the gene function similarity network to generate two enhanced views of the data with edge dropout rates of 0.1, 0.2 and feature masking rates of 0.3, 0.5, respectively, and "Pathway + Pathway" uses the gene pathway co-occurrence network to generate two enhanced views of the data with edge dropout rates of 0.1, 0.2 and feature masking rates of 0.4, 0.5, respectively. Considering that our method consists of multiple parts, different variants were also designed and experimentally compared to validate the contribution of each part. "No adding positive sample pairs" indicates that the positive samples in contrastive learning are only the feature representations of the same genes in different views. "Without network-specific Chebyshev GCN" indicates that only the contrastive learning part is retained without learning the unique feature representations of genes in the network by respective over graph convolution. "Without contrastive learning" indicates no contrastive learning component. "Adding contrastive learning between GO and Pathway" incorporates a comparison between gene functional similarity network and pathway co-occurrence association network into the existing contrastive loss framework. "Predicting by random forest" indicates that random deep forest is used to predict cancer driver genes instead of logistic regression. With the results of the ablation experiments in Table IV, we observed a significant

TABLE IV
ABLATION EXPERIMENTS

Methods	AUC	AUPRC
PPI+PPI	0.9057	0.8264
GO+GO	0.9149	0.8376
Pathway+Pathway	0.9162	0.8300
PPI+GO	0.9230	0.8498
PPI+Pathway	0.9280	0.8522
Pathway+GO	0.9285	0.8561
PPI+Pathway+GO	**0.9317**	**0.8575**
No adding positive sample pairs	0.9287	0.8557
No sharing weight in contrastive learning	0.9281	0.8550
Without network-specific Chebyshev GCN	0.8784	0.7652
Without contrastive learning	0.9192	0.8350
Adding contrastive learning between GO and Pathway	0.9294	0.8553
Predicting by random forest	0.9184	0.8385

The best results are highlighted in bold.

performance degradation when using only a single network for data augmentation and later for contrastive learning, which indicates that using only one network for contrastive provides limited information. The results are significantly improved when using any two networks for contrastive learning. The best performance was achieved when all three networks were used. As the number of networks rises and the number of gene views increases, better results are achieved with multi-view contrastive learning. After removing the strategy of using the neighboring nodes of each gene as positive samples in the contrastive loss, we found a decrease in the model's performance, which indicates that adding positive samples is effective in the contrastive learning encoding module. We also observed that removing the model's contrastive learning part of the model or learning the unique feature part of the gene in the network decreases the model's performance. In our Methods, we conducted contrastive learning between the gene functional similarity network and pathway co-occurrence association network with PPI. We did not perform contrastive learning between GO and Pathway. However, incorporating a comparison of the GO and Pathway networks to the contrastive loss decreases model performance. Because the GO and Pathway networks are generated under different conditions, exhibiting significant differences in the number of edges and nodes between the two networks. Doing a contrast

TABLE V
CO-CITER ANALYSIS OF TOP 30 PAN-CANCER DRIVER GENES PREDICTED BY MNGCL

	#MNGCL	#MRNGCN	#MODIG	#MTGCN	#EMOGI	cancer	driver	biomarker	Drug target	In_NCG	%interact with drivers
TTN	1	271	1	1	1	6	1	4	1	Colorectal Blood Esophagus Brain Stomach	13.43
RYR2	2	360	18	12	18	3	2	1	0	Pleura Breast Blood Colorectal	6.90
FN1	3	26	81	9	13	76	1	16	0		12.11
LRRK2	4	109	584	2	3	16	1	26	6	Kidney Skin Multiple	12.06
SYNE1	5	281	4	38	11	2	1	0	0	Esophagus Breast Bladder head_and_neck Colorectal	19.35
LRP2	6	263	5	3	7	9	1	1	0	Kidney Bladder Brain	11.11
FLG	7	598	3	8	4	6	0	1	1	Multiple Esophagus	16.67
SPTA1	8	77	13	53	9	2	1	1	1	Multiple Brain Uterus Stomach Pancreas	37.93
DST	9	90	16	41	5	4	0	1	0	Brain	8.06
OBSCN	10	228	2	14	109	4	0	0	0	Colorectal	17.65
MACF1	11	17	27	91	19	4	0	0	0	Pancreas Multiple Stomach	18.33
HMCN1	12	717	8	10	48	1	1	0	0	Brain Blood Hepatobiliary Esophagus Colorectal Pancreas	100
APOB	13	36	56	62	30	22	1	12	1	Multiple Hepatobiliary Brain Stomach	15.25
PCLO	14	718	122	5	1214	1	1	0	0	Blood Breast Colorectal	10.0
PRKDC	15	19	532	47	23	268	3	7	4		24.03
RYR1	16	721	215	181	16	3	1	4	0	Kidney	13.04
RYR3	17	948	17	20	71	2	1	0	0		11.11
NEB	18	636	32	11	32	4	1	1	0	Blood Nervous_syste	13.51
FOS	19	9	308	50	351	195	2	9	2		23.94
DMD	20	385	148	2715	40	19	2	10	1	Multiple Colorectal Stomach	3.77
EPHA2	21	4	98	190	262	113	2	4	4	Multiple	31.82
KMT2B	22	43	172	729	1111	3	1	0	0	Multiple Uterus	18.18
USH2A	23	707	237	144	47	4	1	1	0	Breast Uterus Thyroid head_and_neck Pancreas	14.29
HNF4A	24	11	252	126	400	37	1	5	2	Hepatobiliary	31.78
ZFHX4	25	427	11	7	49	3	1	0	0	Lung Kidney head_and_neck	20.00
NOTCH3	26	30	21	437	241	83	1	8	2	Esophagus head_and_neck Multiple Bladder Breast	18.52
FYN	27	12	42	210	21	73	2	0	5	Blood	20.43
APP	28	3	203	4	24	64	6	177	54		7.96
STAT5A	29	61	132	349	384	114	2	5	2		36.19
AHNAK	30	31	813	73	186	10	1	6	1	Kidney	13.98

between them might introduce substantial differences in gene node features that could affect the performance of the contrastive learning encoder. Finally, our model integrates multiple gene networks and uses a contrastive learning strategy to learn the interactions between multiple networks while also considering the unique feature representation of each network to improve the accuracy of cancer driver gene prediction.

F. Performance on Independent Test Set

To verify whether our model is biased to a specific dataset, we also test the MNGCL and all baseline methods on two independent datasets, such as OncoKB and ONGene. We first trained the models with all the pan-cancer positive and negative samples and then used the trained models to predict the genes in OncoKB [44] or ONGene [45] database. OncoKB contains 320 true cancer driver genes, while ONGene contains 388 after removing the genes in the training set. Due to the low number of true positive genes in the test set, all models have low AUPRC values. Fig. 3 shows the AUPRC comparison between our and all baseline methods on the ONGene databases (x-axis) and OncoKB databases (y-axis). The observation shows that our method consistently outperforms the other methods on the two datasets.

G. Predicting New Pan-Cancer Driver Genes

Since the known cancer driver genes are incomplete, we also test the ability of MNGCL to predict new cancer driver genes for pan-cancer. We trained our model with all positive and negative samples base on CPDB PPI network, gene semantic similarity network and pathway co-occurrence association network and applied the model to identify putative driver genes from unmarked genes. It is shown in Table V that the top 30 pan-cancer candidate

driver genes recommended by the MNGCL and their position ranked in other methods (e.g. #MRNGCN is the ranking position in MRNGCN). The number of co-citations between genes and keywords "cancer", "driver", "biomarker", "drug target" are also listed. We checked whether these genes were in the NCG 6.0 list and listed the tissues where these genes locate. We calculated the proportion of our predicted candidate genes related to known driver genes in the PPI network. We observed that 24 of our top 30 predicted genes were recorded as driver genes for at least one cancer type. Moreover, they were all associated with the keyword "cancer". From the experimental results, we also found that all of the 30 new cancer driver candidates we predicted were linked to known cancer driver genes, consistent with the observation that driver genes tend to perform functions in association with each other. Therefore, we believe that our model can find new cancer driver genes.

VI. Conclusion

In this study, we propose a multi-graph contrastive learning approach to predict cancer driver genes called MNGCL. MNGCL constructs three gene-gene networks through protein interaction networks, gene functional similarity, and gene co-occurrence in KEGG pathway. These different networks describe the associations between genes from different perspectives. It then performs contrastive learning among the three networks to obtain interactions between genes from different networks. In addition, considering the information embedded in the networks, it inputs the gene features learned in contrastive learning into three different encoders specific to every network. Finally, we implement the gene feature representations learned from the three networks into a logistic regression model for cancer driver gene prediction. The experimental results of our model show that:

1) MNGCL integrates multiple gene networks and uses a contrastive learning strategy to learn the interactions between multiple networks while also considering the unique feature representation of each network to improve the prediction of cancer driver genes effectively.
2) MNGCL consistently outperformed the baseline approach on both pan-cancer and independent datasets, and achieved the best performance on most single-cancer datasets, demonstrating its superiority in predicting cancer driver genes.
3) The ablation experiments show that we can effectively learn the mutual information between different relational networks through contrastive learning and can help improve the model's performance by further integrating the unique features of each network using logistic regression models.

Our framework is scalable. It can integrate many networks to predict cancer driver genes. We provided experimental results by adding a gene co-expression network into our framework, which led to a limited performance increase (see supplementary files). However, the selected network's reliability and the effectiveness of the data augmentation strategies impact our model's performance. Hence, our future works will employ subgraph sampling methods for data augmentation, such as random walking, to perform contrast learning among different networks without destroying the original network information.

References

[1] B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz Jr., and K. W. Kinzler, "Cancer genome landscapes," *Science*, vol. 339, no. 6127, pp. 1546–1558, 2013.

[2] M. S. Lawrence et al., "Discovery and saturation analysis of cancer genes across 21 tumour types," *Nature*, vol. 505, no. 7484, pp. 495–501, 2014.

[3] J. N. Weinstein et al., "The cancer genome atlas pan-cancer analysis project," *Nature Genet.*, vol. 45, no. 10, pp. 1113–1120, 2013.

[4] J. Zhang et al., "The international cancer genome consortium data portal," *Nature Biotechnol.*, vol. 37, no. 4, pp. 367–369, 2019.

[5] J. G. Tate et al., "Cosmic: The catalogue of somatic mutations in cancer," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D941–D947, 2019.

[6] M. F. Rogers, T. R. Gaunt, and C. Campbell, "Prediction of driver variants in the cancer genome via machine learning methodologies," *Brief. Bioinf.*, vol. 22, no. 4, 2021, Art. no. bbaa250.

[7] R. Andrades and M. Recamonde-Mendoza, "Machine learning methods for prediction of cancer driver genes: A survey paper," *Brief. Bioinf.*, vol. 23, no. 3, 2022, Art. no. bbac062.

[8] X. Shi et al., "Comprehensive evaluation of computational methods for predicting cancer driver genes," *Brief. Bioinf.*, vol. 23, no. 2, 2022, Art. no. bbab548.

[9] J. Wang, X. Chen, Z. Wu, M. Guo, and G. Yu, "Cooperative driver pathways discovery by multiplex network embedding," *Brief. Bioinf.*, vol. 24, no. 3, 2023, Art. no. bbad112.

[10] X. Tang, Q. Xiao, and K. Yu, "Breast cancer candidate gene detection through integration of subcellular localization data with protein–protein interaction networks," *IEEE Trans. Nanobiosci.*, vol. 19, no. 3, pp. 556–561, Jul. 2020.

[11] D. Tamborero, A. Gonzalez-Perez, and N. Lopez-Bigas, "Oncodriveclust: Exploiting the positional clustering of somatic mutations to identify cancer genes," *Bioinformatics*, vol. 29, no. 18, pp. 2238–2244, 2013.

[12] J. Ding et al., "Systematic analysis of somatic mutations impacting gene expression in 12 tumour types," *Nature Commun.*, vol. 6, no. 1, 2015, Art. no. 8554.

[13] M. D. Leiserson et al., "Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes," *Nature Genet.*, vol. 47, no. 2, pp. 106–114, 2015.

[14] A. Cho, J. E. Shim, E. Kim, F. Supek, B. Lehner, and I. Lee, "Muffinn: Cancer gene discovery via network analysis of somatic mutation data," *Genome Biol.*, vol. 17, no. 1, pp. 1–16, 2016.

[15] J. Song, W. Peng, and F. Wang, "An entropy-based method for identifying mutual exclusive driver genes in cancer," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 17, no. 3, pp. 758–768, May/Jun. 2020.

[16] J. Song, W. Peng, F. Wang, and J. Wang, "Identifying driver genes involving gene dysregulated expression, tissue-specific expression and gene-gene network," *BMC Med. Genomic.*, vol. 12, no. 7, pp. 1–12, 2019.

[17] J. P. Hou and J. Ma, "Dawnrank: Discovering personalized driver genes in cancer," *Genome Med.*, vol. 6, pp. 1–16, 2014.

[18] A. C. Gumpinger, K. Lage, H. Horn, and K. Borgwardt, "Prediction of cancer driver genes through network-based moment propagation of mutation scores," *Bioinformatics*, vol. 36, no. Supplement_1, pp. i508–i515, 2020.

[19] A. A. Kumar et al., "pBRIT: Gene prioritization by correlating functional and phenotypic annotations through integrative data fusion," *Bioinformatics*, vol. 34, no. 13, pp. 2254–2262, 2018.

[20] P. Luo, Y. Ding, X. Lei, and F.-X. Wu, "deepDriver: Predicting cancer driver genes based on somatic mutations using deep convolutional neural networks," *Front. Genet.*, vol. 10, 2019, Art. no. 13.

[21] W. Peng, S. Yi, W. Dai, and J. Wang, "Identifying and ranking potential cancer drivers using representation learning on attributed network," *Methods*, vol. 192, pp. 13–24, 2021.

[22] R. Schulte-Sasse, S. Budach, D. Hnisz, and A. Marsico, "Integration of multiomics data with graph convolutional networks to identify new cancer genes and their associated molecular mechanisms," *Nature Mach. Intell.*, vol. 3, no. 6, pp. 513–526, 2021.

[23] W. Peng, Q. Tang, W. Dai, and T. Chen, "Improving cancer driver gene identification using multi-task learning on graph convolutional network," *Brief. Bioinf.*, vol. 23, no. 1, 2022, Art. no. bbab432.

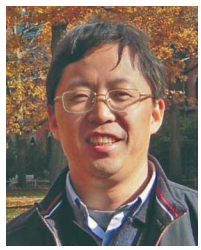
[24] T. Zhang, S.-W. Zhang, M.-Y. Xie, and Y. Li, "A novel heterophilic graph diffusion convolutional network for identifying cancer driver genes," *Brief. Bioinf.*, vol. 24, no. 3, 2023, Art. no. bbad137.
[25] W. Zhao, X. Gu, S. Chen, J. Wu, and Z. Zhou, "MODIG: Integrating multi-omics and multi-dimensional gene network for cancer driver gene identification based on graph attention network model," *Bioinformatics*, vol. 38, no. 21, pp. 4901–4907, 2022.
[26] W. Peng, R. Wu, W. Dai, and N. Yu, "Identifying cancer driver genes based on multi-view heterogeneous graph convolutional network and self-attention mechanism," *BMC Bioinf.*, vol. 24, no. 1, 2023, Art. no. 16.
[27] W. Peng, Z. Che, W. Dai, S. Wei, and W. Lan, "Predicting miRNA-disease associations from miRNA-gene-disease heterogeneous network with multi-relational graph convolutional network model," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 20, no. 6, pp. 3363–3375, Nov./Dec. 2023.
[28] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Representations*, 2017.
[29] X. Wang, N. Liu, H. Han, and C. Shi, "Self-supervised heterogeneous graph neural network with co-contrastive learning," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discov. Data Mining*, 2021, pp. 1726–1736.
[30] C. Wei, J. Liang, D. Liu, and F. Wang, "Contrastive graph structure learning via information bottleneck for recommendation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, pp. 20407–20420.
[31] Y. You, T. Chen, Y. Sui, T. Chen, Z. Wang, and Y. Shen, "Graph contrastive learning with augmentations," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 5812–5823.
[32] Y. Zhu, Y. Xu, F. Yu, Q. Liu, S. Wu, and L. Wang, "Deep graph contrastive representation learning," 2020, *arXiv:2006.04131*.
[33] J. Duan et al., "Graph anomaly detection via multi-scale contrastive learning networks with augmented view," in *Proc. AAAI Conf. Artif. Intell.*, 2023, pp. 7459–7467.
[34] Y. Liu, Y. Zhao, X. Wang, L. Geng, and Z. Xiao, "Multi-scale subgraph contrastive learning," in *Proc. 32nd Int. Joint Conf. Artif. Intell.*, 2023, pp. 2215–2223.
[35] Q. Wang et al., "Unifying cancer and normal RNA sequencing data from different sources," *Sci. Data*, vol. 5, no. 1, pp. 1–8, 2018.
[36] W. E. Johnson, C. Li, and A. Rabinovic, "Adjusting batch effects in microarray expression data using empirical Bayes methods," *Biostatistics*, vol. 8, no. 1, pp. 118–127, 2007.
[37] A. Kamburov, K. Pentchev, H. Galicka, C. Wierling, H. Lehrach, and R. Herwig, "Consensuspathdb: Toward a more complete picture of cell biology," *Nucleic Acids Res.*, vol. 39, no. suppl_1, pp. D712–D717, 2011.
[38] D. Szklarczyk et al., "STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D607–D613, 2019.
[39] D. Repana et al., "The network of cancer genes (NCG): A comprehensive catalogue of known and candidate cancer genes from cancer sequencing screens," *Genome Biol.*, vol. 20, pp. 1–12, 2019.
[40] J. Kim, S. So, H.-J. Lee, J. C. Park, J.-j. Kim, and H. Lee, "Digsee: Disease gene search engine with evidence sentences (version cancer)," *Nucleic Acids Res.*, vol. 41, no. W1, pp. W510–W517, 2013.
[41] G. Yu, F. Li, Y. Qin, X. Bo, Y. Wu, and S. Wang, "GOSemSim: An R package for measuring semantic similarity among go terms and gene products," *Bioinformatics*, vol. 26, no. 7, pp. 976–978, 2010.
[42] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 3844–3852.
[43] D. T. Forster et al., "Bionic: Biological network integration using convolutions," *Nature Methods*, vol. 19, no. 10, pp. 1250–1261, 2022.
[44] D. Chakravarty et al., "Oncokb: A precision oncology knowledge base," *JCO Precis. Oncol.*, vol. 1, pp. 1–16, 2017.
[45] Y. Liu, J. Sun, and M. Zhao, "ONGene: A literature-based database for human oncogenes," *J. Genet. Genomic.*, vol. 44, no. 2, pp. 119–121, 2017.

Wei Peng (Member, IEEE) received the Ph.D. degree in computer science from Central South University, China, in 2013. She is currently a Professor with the Kunming University of Science and Technology, China. Her research interests include bioinformatics and data mining.

Zhengnan Zhou is currently working toward the master's degree with the Kunming University of Science and Technology, China. His research interests include bioinformatics, feature extraction, and data mining.

Wei Dai received the Ph.D. degree in computer application from the University of Chinese Academy of Sciences, China, in 2018. He is currently an Associate Professor with the Kunming University of Science and Technology. His research interests include bioinformatics, distributed and cloud computing, and data mining.

Ning Yu is currently an Associate Professor with the Department of Computing Sciences, State University of New York Brockport, NY, USA. His research interests artificial intelligence, Big Data mining and analysis, and high performance computing.



Jianxin Wang (Senior Member, IEEE) received the B.Eng., M.Eng., and Ph.D. degrees in computer engineering from Central South University, Changsha, China, in 1992, 1996, and 2001, respectively. He is currently the the Dean and a Professor with the School of Computer Science and Engineering, Central South University, Changsha, China. His research interests include algorithm analysis and optimization, parameterized algorithm, bioinformatics and computer network. He has publisher more than 150 papers in various international journals and refereed conferences.