Disease Diagnosis Based on Multi-View Contrastive Learning for Electronic Medical Records

Zhengkang Zhang, Dan Yang, Yu Zhang

***Abstract*—Disease diagnosis based on electronic medical records (EMRs) is one of the important research contents of intelligent healthcare. In recent years, disease diagnosis based on heterogeneous graph neural networks has received increasing attention. However, disease diagnosis tasks often suffer from the lack of labeling information due to the high cost of manual labeling. And existing disease diagnosis models based on heterogeneous graph neural networks ignore the correlation between different meta-paths. Therefore, we propose a disease diagnosis model based on multi-view contrastive learning (MVCDD). MVCDD uses medical data from electronic medical records to construct medical heterogeneous graphs, and uses a fixed-depth random walk method to obtain semantic subgraphs defined by multiple meta-paths. Meanwhile, we introduce the inter-view contrastive learning task to model the correlation between different meta-paths. MVCDD optimizes the patient representations by combining intra-view and inter-view contrastive learning tasks jointly. Extensive experiments are conducted on the MIMIC-III dataset. The experimental results on the MIMIC-III dataset demonstrate that MVCDD outperforms other baselines and effectively improves the performance of disease diagnosis.**

***Index Terms*—Disease Diagnosis, Electronic Medical Records, Graph Neural Networks, Contrastive Learning**

I. Introduction

With the popularization of medical information technology, electronic medical records (EMRs) have received extensive attention. EMRs contains various medical data, such as drugs, procedures, patient vital signs, diagnoses and so on. The abundant medical data in electronic medical records can be utilized for tasks such as disease diagnosis[1] and patient similarity[2,3,4], etc. Many studies have attempted to use EMRs to provide personalized healthcare services to patients. In this paper, we study an important application of EMRs, namely disease diagnosis. The goal of disease diagnosis based on EMRs is to predict the diseases that patients may suffer from based on their basic information in the EMRs, thus assisting medical institutions in diagnosis and improving the efficiency and quality of health care services.

Manuscript received March 10, 2023; revised Jul 31, 2023. This work was supported by the General Scientific Research Project from the Educational Department of Liaoning Province (LJKMZ20220646).

Zhengkang Zhang is a postgraduate student at School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, China (e-mail: aszhangzhengkang@foxmail.com).

Dan Yang is a professor at School of computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, China (corresponding author to provide e-mail: asyangdan@163. com).

Yu Zhang is a senior business manager at China Telecom Digital Intelligence Technology Co., Ltd, Beijing, China (e-mail: zhangy193@chinatelecom.cn).

In recent years, with the advancement of data mining technique, graph neural networks (GNNs) have been increasingly applied to disease diagnosis tasks. However, in practical applications, obtaining manually labeled medical labels is very expensive and time-consuming. Therefore, one of the problems faced by GNNs when applied to the medical field is how to make full use of the unlabeled data on the graph. Recently, inspired by the latest advances in the fields of computer vision and natural language processing, contrastive learning (CL) has been naturally applied to graph learning tasks to address this problem. Graph neural network frameworks based on contrastive learning have great advantages in the absence of labeled information. And they can make full use of the unlabeled data of the graph to learn the embeddings better. For example, the literature [5] proposes cross-view contrastive learning and a view masking mechanism to extract positive and negative samples from two views (network schema view and meta-path view). That enables the two views to collaboratively supervise each other and eventually obtain better embeddings of the nodes.

However, in the real world, most graphs typically exhibit a variety of node and edge types. These types of graphs are commonly known as heterogeneous graphs. As shown in Fig. 1 (a), a medical heterogeneous graph composed of EMRs data contains three types of nodes: patient (P), drug (D), and procedure (O). And it contains two types of edges: patient-drug (representing that the patient takes a certain type of drug) and patient-procedure (representing that the patient has undergone a certain type of procedure). The meta-path is a widely used structure for capturing semantics in heterogeneous graphs. As shown in Fig. 1 (b), there are two meta-paths: patient-drug-patient (PDP) and patient-procedure-patient (POP). The PDP meta-path describes the relationship between patients who have taken the same type of drug and the POP meta-path describes the relationship between patients who have had the same type of procedure. Depending on the selected meta-paths, the relationships between various entities in the medical heterogeneous graph contain diverse semantic information.

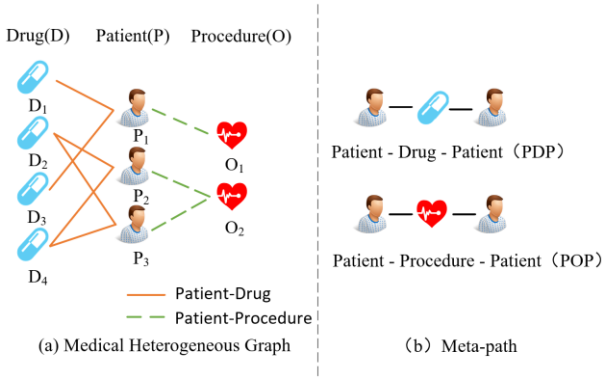

Fig.1 An example of medical heterogeneous graph and meta-path

Currently, disease diagnosis based on heterogeneous graph contrastive learning for EMRs still faces the following problems and challenges:

- **How to effectively capture the rich semantic information in medical heterogeneous graphs to learn a more comprehensive patient representation**. Most existing heterogeneous graph neural network models typically learn embeddings independently from each meta-path, and then integrate them into a final representation. The limitation of this approach is that it ignores the correlation between different meta-paths. For example, in Fig. 1 (a), patient P_2 and patient P_3 are related by having taken the same type of drug D_4, which can be represented by the meta-path PDP. And patients P_2 and P_3 also have a relationship of undergoing the same type of procedure O_2, which can be represented by the meta-path POP. These two meta-paths exhibit evident correlation, and the semantic information derived from one meta-path can assist in learning patient representations from the other meta-path.
- **How to design a effective contrastive learning mechanism for medical heterogeneous graphs for more discriminative embeddings learning**. Medical heterogeneous graphs contain rich semantic information, which is usually reflected by meta-path. Semantic subgraphs defined by different meta-paths contain different semantic information. Contrastive learning only on semantic subgraphs defined by a single meta-path cannot learn embeddings which contain rich semantic information. Therefore, it is particularly important to study the multi-view contrastive learning mechanism.
- **How to efficiently sample semantic subgraphs for medical heterogeneous graphs to learn better embeddings in multi-view contrastive learning**. Electronic medical record contains multiple types of medical data, and the constructed medical heterogeneous graph using electronic medical record has many high-order neighbors for each node. As shown in Fig. 1 (a), patient P_2 and procedure O_2 have a first-order interaction relationship, while patient P_2 and patient P_3 cannot be connected through a first-order interaction relationship. But there exists a second-order interaction relationship between them. Patient P_3 is a high-order neighbor of patient P_2 (i.e., there exists a second-order or higher-order interaction relationship). However, existing graph neural network models use the random walk with restart (RWR) method to sample subgraphs. At each step, the random walk returns to the starting point with probability r. If the RWR method is used to sample semantic subgraphs in the medical heterogeneous graph, it leads to a smaller number of higher-order neighbors of the nodes in the sampled semantic subgraphs. Therefore, a more comprehensive representation of patient representations cannot be learned in multi-view contrastive learning.

To address the above problems and challenges, we propose a disease diagnosis framework based on multi-view contrastive learning MVCDD. The framework first constructs medical heterogeneous graphs using medical data from electronic medical record. And then it samples semantic subgraphs defined by multiple meta-paths using a fixed-depth random walk method. After that, each semantic subgraph is encoded and an inter-view contrastive learning task is introduced to obtain the correlation between different meta-paths. Finally, the patient representations are jointly optimized by combining intra-view and inter-view contrastive learning tasks:

The main contributions are summarized as follows:

- We propose a disease diagnosis framework based on multi-view contrastive learning, MVCDD. The framework uses a fixed-depth random walk method to obtain semantic subgraphs defined by multiple meta-paths. By introducing a multi-view contrastive learning mechanism, our framework can better learn the patient representations using unlabeled data.
- Considering that different meta-paths reflect different semantic information of the same object, and there are correlations between different meta-paths, we introduce the inter-view contrastive learning task to model the correlations between different meta-paths.
- A large number of experiments were conducted on the MIMIC-III dataset. The experimental results show that compared with other mainstream frameworks, the performance of the MVCDD framework proposed in this paper achieves the best.

II. Related Work

This section discusses the work related to intelligent disease diagnosis, including: heterogeneous graph neural networks, contrastive learning, and disease diagnosis based on heterogeneous graph neural networks. Then, the key techniques proposed in this paper are compared with these techniques.

A. Heterogeneous Graph Neural Networks

In recent years, heterogeneous graph neural network has achieved great success in dealing with heterogeneous information networks and has generated widespread interest. For example, HAN [6] proposed the heterogeneous graph attention network that employs a hierarchical attention mechanism to learn node-level and semantic-level structures. Based on HAN, MAGNN [7] additionally considered intermediate nodes of meta-paths. To address the heterogeneity issue concerning edge connections, RSHN [8] built a coarse-grained graph neural network to obtain relational structural features. HetGNN [9] first used random wandering to obtain heterogeneous neighbors of nodes and then used Bi-LSTM to aggregate different types of node features. HGT [10] proposed a heterogeneous graph converter to handle heterogeneous graphs at the network scale.

B. Contrastive Learning

In order to solve the problem that supervised learning paradigm relies on artificial labels, contrastive learning has become a very significant research topic and has been widely applied to graph learning tasks. The contrastive learning model on the graph performs graph learning tasks by distinguishing positive samples and negative samples generated from the graph. DGI [11] compares local

information with global information through Infomax [12] method. On this basis, GMI[13] obtains interaction information separately from node characteristics and topological structures for comparison. MVGRL [14] compares the embeddings from the first-order neighbor and the second-order neighbor. GCC [15] takes the graph as an example and learns to distinguish different instances. GCA [16] randomly deletes the unimportant edges and adds noise to the node characteristics to destroy the attributes. In this way, new views are generated and then the model is optimized by contrastive learning.

C. Disease Diagnosis Based on Heterogeneous Graph Neural Networks

Graph neural networks have become a hot topic in research, as they offer a natural way to model the complex objects and various relationships present in EMRs data. Based on this, many researchers have started to use graph neural networks for disease diagnosis tasks. For example, a disease diagnosis model based on graph neural networks is proposed in the literature [17]. Considering the problem of insufficient data volume of electronic medical records (EMRs), the model additionally introduces a medical knowledge base. The medical concept graph is constructed based on the external medical knowledge base, while the patient record graph is constructed using EMRs. Then the embeddings of patient nodes and disease nodes are learned by aggregating the information of direct neighbor nodes through graph encoders for the disease prediction task. Considering the correlation between various modalities, a multimodal learning framework is proposed in the literature [18]. The attention mechanism is used to capture the correlation and complementarity between modalities for the disease diagnosis task.

The disease diagnosis framework based on multi-view contrastive learning proposed in this paper differs from the above studies in the following ways:

- The existing disease diagnosis models based on heterogeneous graph neural networks overlook the correlation between different meta-paths. In contrast, the framework proposed in this paper introduces an inter-view contrastive learning task to model the correlation between different meta-paths, thereby learning more comprehensive patient representations.
- Most of the existing contrastive learning models on graphs focus on homogeneous graphs and cannot handle the rich semantic information on heterogeneous graphs. The framework in this paper concentrates on heterogeneous graphs and compares multiple semantic subgraphs defined by different meta-paths, which can better handle the rich semantic information in heterogeneous graphs and improve the performance of the framework.
- Most existing disease diagnosis models use a supervised learning paradigm and rely on manually annotated data. However, the cost of manually labeling EMRs data is very expensive. In contrast, the framework in this paper uses a contrastive learning method, which can effectively learn the patient representations better using unlabeled data.

This paper combines the three advantages mentioned above and propose a disease diagnosis framework based on multi-view contrastive learning. The framework constructs medical heterogeneous graphs using the patient's electronic medical record, and performs contrastive learning on multiple semantic subgraphs defined by different meta-paths. It combines intra-view and inter-view contrastive learning tasks to jointly optimize the patient representations.

III. PRELIMINARIES

In this section, we formally define some key concepts in the disease diagnosis framework as follows.

Definition 1. Medical heterogeneous graph. A medical heterogeneous graph is defined as G = (V, E, A, R). It associates a node-type mapping function Φ: V→A and an edge-type mapping function Ψ: E→R. A and R denote the sets of node-type and edge-type respectively, and | A | + | R | >2..

Definition 2. Meta-path. A meta-path P is defined as a path. It has the form $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \cdots \xrightarrow{R_l} A_{l+1}$(Abbreviations are $A_1A_2 \ldots A_{l+1}$), describes a composite relationship between node types A_1 and A_{l+1} : $= R_1 \circ R_2 \circ \cdots \circ R_l$, where $\circ$ denotes the complex operator on the relation

The symbols commonly used in this paper are specified as shown in Table I.

TABLE I
SYMBOLS AND THEIR DEFINITIONS

Symbol	DEFINITION
G	Medical Heterogeneous Graph
V,E	Node and edge sets
A,R	Node type and edge type sets
P	Meta-path
G_p	semantic subgraph set
z_i^p	Embeddings of the node in the semantic subgraph
$\hat{z}_i^{s,t}$	Semantic embeddings of the target semantic subgraph P_t decoded from the source semantic subgraph P_s

IV. DISEASE DIAGNOSIS FRAMEWORK

This section describes the proposed disease diagnosis framework MVCDD in detail. The overall framework structure is shown in Fig. 2, where medical heterogeneous graphs are constructed using medical data in EMR, and semantic subgraphs defined by multiple meta-paths are sampled using a fixed-depth random wandering approach to improve disease diagnosis performance by combining intra-view and inter-view contrastive learning tasks to jointly optimize the patient representations.

A. Constructing Medical Heterogeneous Graph

For constructing medical heterogeneous graph, the objective is to utilize multiple types of medical data in electronic medical records to improve the performance of the framework for disease diagnosis. In this paper, we construct a medical heterogeneous graph by extracting multiple types of medical data of patients from EMR data. As an example, the medical heterogeneous graph in Fig. 1 (a) contains three types of nodes: patient (P), drug (D), and procedure (O), and two types of edges: patient-drug (the patient took the type of drug), and patient-procedure (the

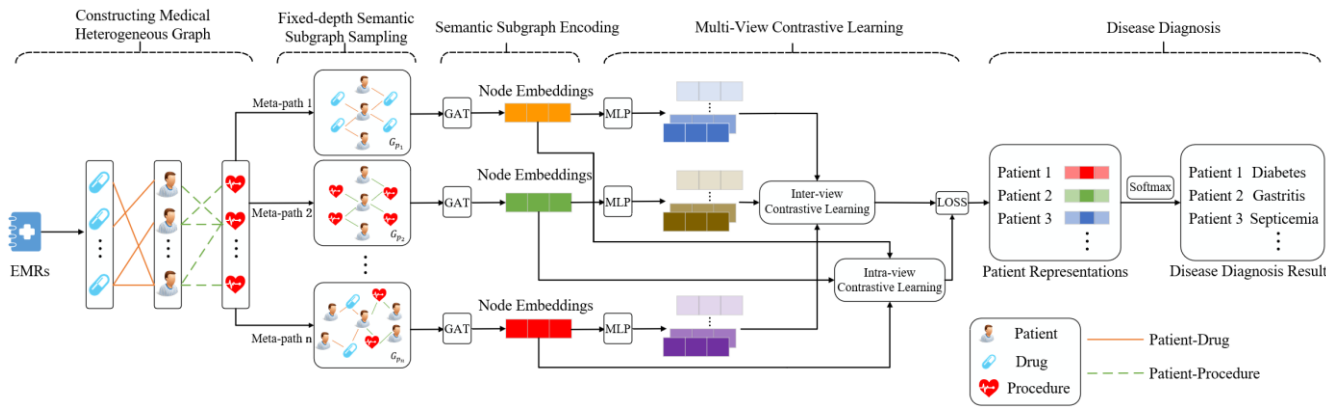

Fig.2 The overall architecture of MVCDD

patient had the type of procedure). By using three sets of node types (patients, drugs, procedures) and different edge relationships between nodes to connect them, a medical heterogeneous graph is constructed. Given a medical heterogeneous graph G=(V, E, A, R), drug node V_{d_l}={V_{d_1}, V_{d_2}, ...,V_{d_l}}, l is the number of drugs, procedure node V_{o_m}={V_{o_1}, V_{o_2}, ...,V_{o_m}}, m is the number of procedures, and the edge E_r={E_1, E_2, ...,E_r}, r is the number of edges. Considering the adjacency matrix H of the medical heterogeneous graph G, if patient i has taken drug V_{d_l} or undergone procedure V_{o_m}, then the drug node V_{d_l} or the procedure node V_{o_m} belongs to edge E_r .Then the corresponding position of the adjacency matrix H is set to 1, otherwise it is set to 0.

B. Fixed-depth Semantic Subgraph Sampling

The original graph neural network model [15,19] uses the random walk with restart (RWR) method to sample the subgraph. At each step, the random walk returns to the starting point with probability r. However, there are more high-order neighbors of nodes in medical heterogeneous graphs which constructed by using patients' electronic medical records. If the RWR method is used to sample semantic subgraphs of medical heterogeneous graphs, it leads to fewer higher-order neighbors of nodes in semantic subgraphs, which impairs the rich semantic information in medical heterogeneous graphs.

To address the problem mentioned, a fixed-depth random walk method is employed for semantic subgraph sampling. Specifically, this method performs a random walk guided by the meta-path P and samples semantic subgraphs with probability proportional to the edge weights of the meta-path constraint relations. When the depth of the random walk reaches the maximum depth K_P set, the random walk stops. The semantic subgraph set G_p= {G_p, $P \in M$} consists of all nodes sampled in n random walks. Since the maximum depth of the random walks is fixed, more high-order neighbors of patient nodes can be obtained by specifying the maximum depth K_P to learn a better patient representation.

Given a node v_i in a medical heterogeneous graph G and a meta-path P from the meta-path set M = {M_1, M_2, ..., M_t}, MVCDD obtains the semantic subgraph set G_p= {G_p, $P \in M$} by a fixed-depth random walk method. After that, the adjacency matrix and node feature matrix of each semantic subgraph are generated as the input of the next module.

For the patient nodes feature matrix, we use various types of medical data from electronic medical records as patient features, such as drugs, procedures, etc. We use one-hot to encode the characteristics of patients.

C. Semantic Subgraph Encoding

Electronic medical records contain many types of medical data of patients, and different types of medical data have different degrees of association with the diseases obtained by patients. To capture the degree of impact that different types of medical data have on patients, we use GAT [20] as a graph encoder. MVCDD assigns different weights to nodes of different types in the semantic subgraph by using attention mechanism. Given a semantic subgraph G_p= {G_p, $P \in M$} defined by a meta-path P, node v_i and node $v_j \in N_j^p$, where N_j^p is the neighborhood node of node v_i in the semantic subgraph defined by meta-path P. First, the attention coefficient $a_{i,j}$ between nodes v_i and v_j in each semantic subgraph is calculated with the following formula:

$$a_{i,j} = \frac{\exp\left(LeakyReLU(a^T[Wh_i \| Wh_j])\right)}{\sum_{l \in N_j} \exp\left(LeakyReLU(a^T[Wh_i \| Wh_l])\right)} \quad (1)$$

Where a^T denotes the transpose of the weight vector, W is the trainable weight matrix, h_i, h_j, h_l are the feature matrices of the corresponding nodes, $\|$ is the splicing operation. After that, we use a multi-head attention mechanism to obtain the embeddings of the semantic subgraph z_i^p = {$z_i^p \in R^{1 \times d_s}$, $P \in$ M }, The formula is as follows:

$$z_i^P = \|_{K=1}^{K} \sigma\left(\sum_{j \in N_i^p} a_{ij}^k W^K z_j^p\right) \quad (2)$$

Where $\|$ is the splicing operation, K represents the total number of heads of the multi-head attention mechanism, σ is the activation function, N_i^p denotes the neighbors of node v_i in the semantic subgraph set G_p, a_{ij}^k is the normalized attention coefficient computed at the K-th head, and W^K is the weight matrix of the K-th head.

D. Multi-View Contrastive Learning

For multi-view contrastive learning, the goal is to obtain an accurate patient representation by combining intra-view and inter-view contrastive learning, where the views in this process are the semantic subgraphs.

1) Intra-view Contrastive Learning

We use MoCo [21] as the contrastive learning framework

to maintain the dictionary as a negative sample queue and encode new keys instantaneously by a momentum updating encoder. Intra-view contrastive learning selects positive and negative samples from each semantic subgraph itself for comparison. In each semantic subgraph, the nodes that have direct interaction with the target node are considered as positive samples, while the remaining nodes are considered as negative samples. In each epoch of MVCDD, the query node is compared with all the negative sample nodes in the dictionary. Multi-View contrastive learning is performed separately for each semantic subgraph by InfoNCE [22] loss, and the intra-view contrast loss function is as follows

$$\mathcal{L}_{intra} = \frac{1}{|M|}\sum_{P\epsilon M} -log\frac{exp\,(z_i^P \cdot z_{K_+}^P/\tau)}{\sum_{j=0}^{K} exp\,(z_i^P \cdot z_{K_j}^P/\tau)} \quad (3)$$

Where z_i^P is the embedding of the node in the semantic subgraph defined by the meta-path P calculated by Equation 2, $z_{K_+}^P$ represents the positive samples matching the query node, $z_{K_j}^P$ represents the negative samples, K is the size of the dictionary, and τ is the temperature hyper-parameter. The value of $\mathcal{L}_{intra}$ is lower when the query node z_i^P is similar to its positive sample $z_{K_+}^P$ but not similar to all other nodes (which are considered as negative samples of z_i^P). Therefore, by minimizing $\mathcal{L}_{intra}$, the MVCDD framework is able to use each meta-path in M to distinguish the subgraph instances of different nodes.

2) Inter-view Contrastive Learning

The previous models of heterogeneous graph neural network independently learn embeddings from each meta-path and finally integrates them into the final representation. However, different meta-paths reflect different semantic information of the same object, and there are correlations between meta-paths. If the correlation between meta-paths is ignored in contrastive learning, the rich semantic information in the semantic subgraphs cannot be captured. Therefore, we introduce the inter-view contrastive learning task to capture the correlation between meta-paths. For each semantic embedding z_i^P of node v_i in the semantic subgraph defined by meta-path P, MVCDD decodes it into semantic embeddings of other semantic subgraphs to preserve the correlation between meta-paths. In this paper, MLP is chosen as the decoder with the following formula:

$$\hat{z}_i^{s,t} = g^{s,t}(z_i^{P_s}) \quad (4)$$

Where $g^{s,t}()$ represents the decoder that decodes the semantic embeddings from the source semantic subgraph P_s to the target semantic subgraph P_t. $\hat{z}_i^{s,t}$ represents the semantic embeddings of the target semantic subgraph P_t decoded from the source semantic subgraph P_s. For example, if the meta-path of the source semantic subgraph is set to PDP and the meta-path of the target semantic subgraph is set to POP, the decoder tries to use the relationship of whether the patients have taken the same type of drug among themselves to predict the relationship of whether the patients have had the same type of procedure among themselves. In this way, MVCDD captures the correlation between different meta-paths. The inter-view contrastive learning takes the nodes in the target semantic subgraph that have direct interaction with the target node as positive samples, and the other nodes in the target semantic subgraph as negative samples. The loss function of inter-view contrastive learning is as follows:

$$\mathcal{L}_{inter} = \frac{1}{|M|*(|M|-1)}\sum_{P_s,P_t\epsilon M, s\neq t} -log\frac{exp\,(\hat{z}_i^{s,t} \cdot z_{K_+}^{p_t}/\tau)}{\sum_{j=0}^{K} exp\,(\hat{z}_i^{s,t} \cdot z_{K_j}^{p_t}/\tau)} \quad (5)$$

Where $|M|$ represents the number of meta-paths, $z_{K_+}^{p_t}$ represents the positive samples matching the query nodes, and $z_{K_j}^{p_t}$ represents the negative samples. Finally, the MVCDD framework optimizes the overall loss $\mathcal{L}$ jointly by combining intra-view and inter-view contrastive learning tasks to fully learn the patient representations:

$$\mathcal{L} = \alpha\mathcal{L}_{intra} + (1-\alpha)\mathcal{L}_{inter} \quad (6)$$

Where α is the hyper-parameter used to balance the different loss functions.

E. Disease Diagnosis

The final patient representations h_i is obtained by the above calculation. The predicted values of different disease labels are obtained by classifying the final patient representations h_i using the Softmax function. The disease label with the highest prediction value is taken as the patient's disease diagnosis result $\hat{y}$. The calculation formula is as follows

$$\hat{y} = Softmax(h_i) \quad (7)$$

V. Experiments And Evaluation

This section first introduces the dataset and preprocessing, evaluation metrics and baselines used for the experiments, and then illustrate the performance of MVCDD based on the experimental data.

A. Dataset and Preprocessing

We utilize the real electronic medical records dataset, MIMIC-III [23]. Two types of information about the patient are used to construct the medical heterogeneous graph for disease diagnosis tasks, i.e., drugs and procedure. In this paper, we preprocessed data from the MIMIC-III dataset and select six disease labels: Coronary Disease, Diabetes, Heart Failure, Gastritis, Respiratory Failure and Septicemia. First, patients were screened to remove patients with missing medication or surgical information. For each patient, only drugs of type 'MAIN' were selected. The drugs of type 'BASE' and drugs used less than M times were removed, where M is the hyper-parameter. After that, the drugs taken by each patient were randomly selected, and the number of drugs used by each patient was controlled to be within 30. For each patient's procedures, the procedures were ranked according to their importance, and the number of procedures was controlled to be within 10. Finally, the information of 7736 patients was obtained, and the statistics of the processed data set are shown in Table II.

TABLE II
STATISTICS OF DATASETS

Disease label	Number of patients
Coronary Disease	3017
Diabetes	224
Heart Failure	855
Gastritis	415
Respiratory Failure	1288
Septicemia	1937
Total	7736

TABLE III
RESULTS OF DISEASE DIAGNOSIS EXPERIMENTS USING DIFFERENT METHODS

Methods	Percentage of labeled nodes =5%		Percentage of labeled nodes =10%		Percentage of labeled nodes =15%	
	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1
GCN	66.14	80.19	69.71	81.68	71.79	82.14
GAT	67.46	81.49	70.36	83.08	74.75	83.29
HAN	62.55	79.30	64.30	80.47	67.97	82.55
Heco	69.28	80.34	73.98	82.69	75.47	83.37
MVCDD-GIN	68.66	80.11	73.65	81.60	75.83	83.99
MVCDD-GCN	72.61	81.73	75.29	83.51	76.55	83.61
MVCDD	72.91	82.07	76.30	84.03	78.42	84.73

B. Evaluation Metrics

We used Macro-F1 and Micro-F1 as evaluation metrics for disease diagnostic tasks:

1) Macro-F1

Macro-F1 score calculates the arithmetic average of F1 scores for all categories, and is not affected by data imbalance. It is suitable for multi-classification scenarios:

$$Macro - F1 = \frac{1}{n}\sum_{i=1}^{n}\frac{2TP_i}{(2TP_i + FP_i + FN_i)} \tag{8}$$

2) Micro-F1

Micro-F1 score calculates the weighted average of F1 scores of all categories, and is susceptible to data imbalance. It is suitable for multi-classification scenarios:

$$Micro - F1 = \frac{\sum_{i=1}^{n} 2TP_i}{\sum_{i=1}^{n}(2TP_i + FP_i + FN_i)} \tag{9}$$

Where n represents the number of disease categories, TP_i, FP_i, FN_i respectively denote the number of true positive, false positive and false negative in the class i disease.

C. Baselines

To test the performance of MVCDD framework, we compare MVCDD with the following baseline methods:

- GCN[24]. This method based on convolutional neural networks and is widely used for graph embedding.
- GAT[20]. This method Introduces a self-attention mechanism to aggregate features by calculating the attention coefficients of nodes.
- HAN[6]. This method uses a hierarchical attention mechanism to learn node-level and semantic-level structures
- HeCo[5]. This method extracts positive and negative embeddings from the network schema view and meta-path view, and learns the final node embeddings using a cross-view contrastive learning mechanism.
- MVSDD-GIN. This framework is a variant of MVCDD. When using this method in comparison experiments, the encoder uses the GIN[25].
- MVSDD-GCN. This framework is a variant of MVCDD. When using this method in comparison experiments, the encoder uses the GCN.

D. Implementation Details

For GCN, GAT, HAN, HeCo, the parameters follow the settings in the original paper and the best performance is reported. For MVCDD-GIN, MVCDD-GCN, the encoder follows the parameter settings of the original paper, and the other parameters are consistent with the MVCDD framework.

For the MVCDD framework proposed in this article, at each epoch, 32 fixed-depth random walks are performed to sample semantic subgraphs constrained by element paths, where the maximum depth is set to twice the depth of the element path, i.e., K_P=2 | P |, where | P | is the depth of the element path. A 2-layer MLP is used as a decoder to model the correlation between different element paths. The Adam [26] optimizer is used during training, with a learning rate of 0.003, semantic subgraph embedding dimensions set to 64, and node embedding dimensions d set to 64. We selected three meta-paths: patient-drug-patient (PDP), patient-procedure-patient (POP) and patient-drug-patient-procedure-patient (PDPOP). During the preprocessing of the dataset, drugs that were used less than M times were removed, and M was set to 50.

For MoCo [21], the dictionary size K is set to 7740 and temperature hyper-parameter τ = 0.07. For the fine-tuning phase, the training is performed using a 2-layer GCN [11], and the weight decay is set to 1e-5.

E. Experimental Results and Analysis

The performance of MVCDD was tested by completing a disease diagnosis task in the experiment. Given the medical data of patients in the EHR (e.g., drugs the patients have taken, procedures the patient have undergone, etc.), the disease diagnosis task is to predict the diseases that patients may have. Unlabeled patient nodes were used for training during the pre-training process of MVCDD. During the fine-tuning process, 5%, 10%, and 15% of labeled patient nodes were randomly selected as training sets, and the remaining labeled nodes were used as the test set.

The experimental results are shown in Table III. Based on the results, the following conclusions can be drawn:

- The performance of the proposed MVCDD framework in this paper consistently outperforms other baseline methods. It is shown that by additionally introducing an inter-view contrastive learning task to capture the correlation between meta-paths, the patient representations can be effectively optimized and the disease diagnosis performance of the framework can be improved.
- Methods based on contrastive learning (HeCo, MVCDD-GIN, MVCDD-GCN, MVCDD) exhibit

better performance than other methods. This suggests that the pre-training process of contrastive learning methods can use a large amount of unlabeled data to obtain embeddings with rich semantic information, which makes them perform better in disease diagnosis tasks. When the proportion of labeled nodes is low, the performance improvement of contrastive learning methods for disease diagnosis is more remarkable.

F. Variant Analysis

To validate the rationality of the MVCDD framework structure, this paper proposes two variations of MVCDD: MVCDD-Intra and MVCDD-Inter. MVCDD-Intra only considers the intra-view contrastive learning task, while MVCDD-Inter only considers the inter-view contrastive learning task. The performance of the two variations is compared with MVCDD on the MIMIC-III dataset. The experimental results use Macro-F1 and Micro-F1 as evaluation metrics, as shown in Fig. 3.

From this, we can draw the following conclusions:

- The results of MVCDD consistently outperform other variants, demonstrating the effectiveness and necessity of combining intra-view and inter-view contrastive learning tasks for multi-view contrastive learning.
- The results of MVCDD-Inter outperform the results of MVCDD-Intra, indicating the importance of introducing an inter-view contrastive learning task that captures correlations between meta-paths.

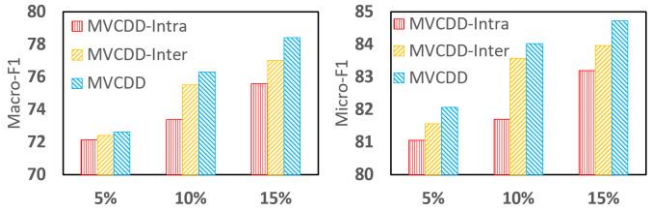

Fig.3 The comparison of MVCDD and its variants

G. Visualization

To visually evaluate the disease diagnosis results, we use t-SNE [27] to map the patient nodes embeddings in the test set to a two-dimensional space. The results of the visualization experiments are shown in Fig. 4, where different colors represent different types of disease labels.

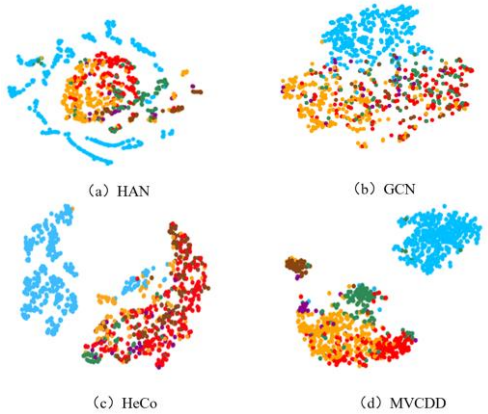

Fig.4 Visualization of the patient nodes embeddings

It can be seen that the differently labeled patient nodes in HAN are not well aggregated and the differently labeled patient nodes in GCN are not well separated. The contrastive learning methods (Heco, MVCDD) produce more distinct boundaries and fewer overlapping regions. For HeCo, the nodes are still mixed to some extent because it fails to capture the correlation between meta-paths. The MVCDD proposed in this paper properly separates patient nodes with different labels with relatively clear boundaries, and patient nodes with the same labels are also better aggregated together, indicating that better patient nodes embeddings are learned.

H. Parameter Analysis

This section discusses the parameter impact analysis of MVCDD, which mainly involves four important hyper-parameters: the number of fixed-depth random walks N, the dictionary size K, the node embedding dimension d and meta-path. Taking 15% of the labeled nodes as the training set, we change the values of N, K and d by keeping the other parameters constant and observe the performance change of MVCDD. Macro-F1 and Micro-F1 are used as evaluation metrics. Fig. 5 - Fig. 7 respectively show the performance variation of MVCDD under different fixed-depth of random walks, dictionary sizes, and node embedding dimensions. Table IV shows the performance of different meta-paths.

a) Different fixed-depth of random walks

The number of fixed-depth random walks determines the number of higher-order neighbors of the patient nodes in the sampled semantic subgraph. As shown in Fig. 5, the performance of MVCDD initially improves and then declines as the number of random walks N increases. The best result is achieved when the number of random walks is set to 32. The reason for this phenomenon is that when the number of fixed-depth random walks is too large, nodes that are not related to the patient node are also sampled into the semantic subgraph, which is not conducive to learning the embeddings of patient nodes.

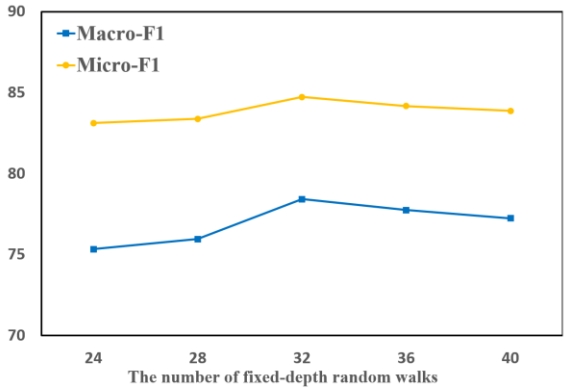

Fig.5 Performance variation of MVCDD with different number of fixed-depth random walk

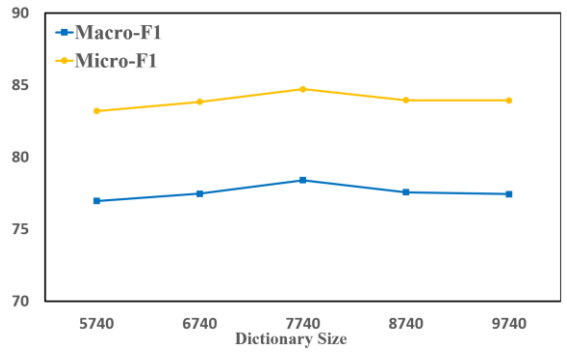

Fig.6 Performance variation of MVCDD with different dictionary sizes

b) Dictionary sizes

The dictionary size determines the number of negative samples in multi-view contrastive learning. As shown in Fig. 6, when the dictionary size is set to 7740, the performance of MVCDD is the best. And the performance gradually improves when the dictionary size is smaller than 7740, and decreases when the dictionary size is larger than 7740.

c) Node embedding dimensions

As shown in Fig. 7, the performance is optimal when the node embedding dimension is set to 64. The performance decreases gradually when the value is greater or less than 0.5.

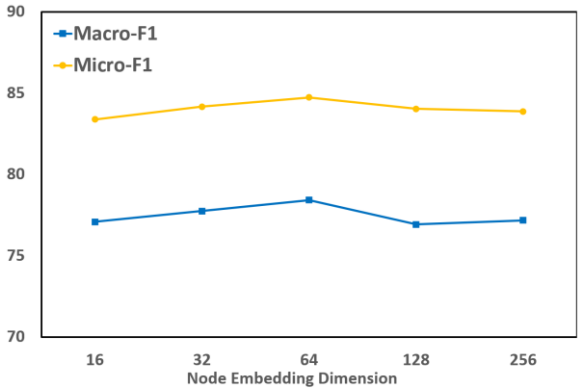

Fig.7 Performance variation of MVCDD with different node embedding dimensions

d) Meta-path

MVSDD obtains different semantic subgraphs according to different meta-paths, and different semantic subgraphs capture different semantic information in the medical heterogeneous graph. As shown in Table IV, the effect of selecting three meta-paths PDP, POP and PDPOP is significantly better than the effect of arbitrarily selecting two meta-paths, reflecting the effectiveness of selecting semantic subgraphs defined by multiple meta-paths for contrastive learning.

TABLE IV
THE PERFORMANCE OF DIFFERENT META-PATHS

Meta-path	Macro-F1	Micro-F1
PDP，POP	75.49	82.79
PDP, PDPOP	73.18	81.17
POP, PDPOP	75.55	83.77
PDP, POP, PDPOP	78.42	84.73

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a disease diagnosis framework based on multi-view contrastive learning (MVCDD). The framework constructs a medical heterogeneous graph using electronic medical records and introduces a multi-view contrastive learning mechanism to improve disease diagnosis performance. When MVCDD performs semantic subgraph sampling, considering the large number of high-order neighbors of patient nodes in medical heterogeneous graphs constructed from electronic medical record data, a fixed-depth random walk method is used to obtain semantic subgraphs defined by multiple meta-paths. We introduce an inter-view contrastive learning task to model the correlation between different meta-paths, and combines intra-view and inter-view contrastive learning tasks to jointly optimize the framework. Extensive experiments were conducted on the MIMIC-III dataset, and the experimental results showed that MVCDD outperformed baseline methods and could effectively utilize unlabeled data to learn better patient representations.

In the following work, we will investigate how to use multiple modalities of medical data in electronic medical records, such as patient chest X-rays. The fusion of multimodal features can help to obtain more accurate patient representations and thus improve disease diagnosis performance.

REFERENCES

[1] Rushan Long, Dan Yang, and Yang Liu. DiseaseNet: A Novel Disease Diagnosis Deep Framework via Fusing Medical Record Summarization[J]. IAENG International Journal of Computer Science, vol. 49, no.3, pp808-817, 2022.

[2] Yi Li, Dan Yang, and Xi Gong. Patient Similarity via Medical Attributed Heterogeneous Graph Convolutional Network[J]. IAENG International Journal of Computer Science, vol. 49, no.4, pp1152-1161, 2022.

[3] Zhihuang Lin, and Dan Yang. Medical Concept Embedding with Variable Temporal Scopes for Patient Similarity[J]. Engineering Letters, vol. 28, no.3, pp651-662, 2020

[4] Hua Jiang, and Dan Yang. Learning Graph-based Embedding from EHRs for Time-aware Patient Similarity[J]. Engineering Letters, vol. 28, no.4, pp1254-1262, 2020

[5] Wang Xiao, Liu Nian, Han Hui, et al. Self-supervised Heterogeneous Graph Neural Network with Co-contrastive Learning [C]. //Proc of the 27th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. pp1726-1736, 2021.

[6] Wang Xiao, Ji Houye, Shi Chuan, et al. Heterogeneous Graph Attention Network[C].//Proc of the 28th Int Conf on World Wide Web. pp2022-2032, 2019.

[7] Fu Xinyu, Zhang Jiani , Meng Ziqiao, et al. MAGNN: Metapath Aggregated Graph Neural Network for Heterogeneous Graph Embedding[C]//Proc of the 29th Int Conf on World Wide Web. pp 2331-2341, 2020.

[8] Zhu Shichao, Zhou Chuan, Pan Shirui, et al. Relation Structure-Aware Heterogeneous Graph Neural Network[C]//Proc of the 19th IEEE International Conference on Data Mining. pp1534-1539, 2019.

[9] Zhang Chuxu, Song Dongjin, Huang Chao, et al. Heterogeneous Graph Neural Network[C].//Proc of the 25th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. pp793–803, 2019.

[10] Hu Ziniu, Dong Yuxiao, Wang Kuansan, et al. Heterogeneous Graph Transformer[C].//Proc of the World Wide Web Conference. pp2704-2710, 2020.

[11] Petar Velickovic, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, and R. Devon Hjelm. Deep Graph Infomax. In ICLR. 2019.

[12] Ralph Linsker. Self-Organization in a Perceptual Network. Computer, pp105–117, 1988.

[13] Zhen Peng, Wenbing Huang, Minnan Luo, et al. Graph Representation Learning via Graphical Mutual Information Maximization. In WWW, pp259–270, 2020.

[14] Kaveh Hassani, Amir Hosein Khasahmadi. Contrastive Multi-View Representation Learning on Graphs[C].//Proc of the 37th International Conf on Machine Learning. pp4116–4126, 2020.

[15] Qiu Jiezhong, Chen Qibin, Dong Yuxiao, et al. GCC: graph contrastive coding for graph neural network pre-training[C]. //Proc of the 26th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining.pp1150–1160, 2020.

[16] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Graph Contrastive Learning with Adaptive Augmentation. In Proceedings of the Web Conference. pp2069–2080, 2021.

[17] Sun zhenchao, Yin Hongzhi, Chen Hongxu, et al. Disease Prediction via Graph Neural Networks[J]. IEEE Journal of Medical Biomedical and Health Infomatics. pp818-826, 2021.

[18] Zheng Shuai, Zhu Zhenfeng, Liu Zhizhe, et al. Multi-modal Graph Learning for Disease Prediction[J]. IEEE Transactions on Medical Imaging. pp2207-2216, 2022.
[19] Dong Yuxiao, V. Chawla Nitesh, Swami Ananthram, et al. metapath2vec: Scalable Representation Learning for Heterogeneous Networks [C]. //Proc of the 23th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. pp135–144, 2017.
[20] Petar Veličković, Guillem Cucurull, Arantxa Casanova, et al. Graph attention networks [C]. //International Conference on Learning Representations. 2018.
[21] He Kaiming, Fan Haoqi, Wu Yuxin, et al. Momentum Contrast for Unsupervised Visual Representation Learning[C]//Proc of the IEEE Conference on Computer Vision and Pattern Recognition.pp9726–9735, 2020.
[22] Aaron van den Oord, Yazhe Li, Oriol Vinyals. Representation Learning with Contrastive Predictive Coding[J]. arXiv e-prints, 2018.
[23] A. E. Johnson, T. J. Pollard, L. Shen, et al. MIMIC-III, a freely accessible critical care database, Scientific data, vol. 3, no. 1, pp1–9, 2016.
[24] Thomas N. Kipf, Max Welling. Semi-supervised classification with graph convolutional networks[C]. International Conference on Learning Representations. 2017.
[25] Xu Keyulu, Hu Weihua, Leskovec Jure, et al. How powerful are graph neural networks?[C]. //International Conference on Learning Representations. 2019.
[26] Diederik P. Kingma, Jimmy Lei Ba. Adam: A method for stochastic optimization[C]. //International Conference on Learning Representations. 2015.
[27] Laurens van der Maaten, Geoffrey Hinton. Visualizing Data using t-SNE [J]. Journal of Machine Learning Research. pp2579 – 2605, 2008.