

Article

DRR: Global Context-Aware Neural Network Using Disease Relationship Reasoning and Attention-Based Feature Fusion

Zhixing Ding [†], Zhengqiang Li [†], Xi Li and Hao Li ^{*}

The School of Information Science and Engineering, Yunnan University, Kunming 650504, China; dzxdwd@mail.ynu.edu.cn (Z.D.); lizhengqiang@stu.ynu.edu.cn (Z.L.); lixi103@mail.ynu.edu.cn (X.L.)

^{*} Correspondence: lihao707@ynu.edu.cn

[†] These authors contributed equally to this work.

Abstract: The prediction of future disease development based on past diagnosis records has gained significant attention due to the growing health awareness among individuals. Recent deep learning-based methods have successfully predicted disease development by establishing relationships for each diagnosis record and extracting features from a patient's past diagnoses in chronological order. However, most of these models have ignored the connections between identified diseases and low-risk diseases, leading to bottlenecks and limitations. In addition, the extraction of temporal characteristics is also hindered by the problem of global feature forgetting. To address these issues, we propose a global context-aware net using disease relationship reasoning and attention-based feature fusion, abbreviated as DRR. Our model incorporates a disease relationship reasoning module that enhances the model's attention to the relationship between confirmed diseases and low-risk diseases, thereby alleviating the current model's bottlenecks. Moreover, we have established a global graph-based feature fusion module that integrates global graph-based features with temporal features, mitigating the issue of global feature forgetting. Extensive experiments were conducted on two publicly available datasets, and the experiments show that our method achieves advanced performance.

Keywords: disease prediction; disease relationship reasoning; graph-based feature

MSC: 68T07



Citation: Ding, Z.; Li, Z.; Li, X.; Li, H. DRR: Global Context-Aware Neural Network Using Disease Relationship Reasoning and Attention-Based Feature Fusion. *Mathematics* **2024**, *12*, 488. <https://doi.org/10.3390/math12030488>

Academic Editor: Dechao Bu

Received: 27 December 2023

Revised: 20 January 2024

Accepted: 29 January 2024

Published: 2 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The growing interest in using artificial intelligence (AI) [1] to improve healthcare delivery and patient outcomes through electronic health records (EHRs) reflects a shift towards leveraging technology in medicine, especially in predicting patient health outcomes. A vital component of this progression is the extraction and utilization of patient characteristics from longitudinal EHRs [2,3]. The more detailed the patient data available, the more sophisticated and effective the resulting medical AI systems can be. In recent years, there has been a significant surge in the adoption of Electronic Health Record (EHR) systems globally, leading to the accumulation of substantial amounts of electronic patient data. These data include both structured components, such as disease and medication codes, and unstructured elements, such as clinical narratives and progress notes.

Given that EHRs consist of both structured and unstructured data, the application of deep learning techniques is an excellent approach. Deep learning can effectively process and interpret the varied and complex layers of data in EHRs. This ability allows for the extraction of a wide range of information, which can be used to predict the likelihood of a patient developing certain medical conditions, anticipate reactions to specific medications, and forecast future health conditions. The use of deep learning for disease prediction in the context of EHRs represents a forward-thinking strategy in leveraging the rich, multifaceted data available in these records. In this realm, disease prediction models using deep learning are primarily categorized into three types: models based on the Transformer

architecture [4–6], time-series models based on recurrent neural networks (RNNs) [7,8], and models based on convolutional neural networks (CNNs) [9] and graph-based [10–13]. One of the main keys behind these models for disease prediction tasks is to mine information from patients' previous diagnostic data at each diagnosis and utilize it to forecast patients' future disease progression. How models learn disease features, especially the contextual relationships between diseases, remains a challenge to this day.

Recently, Lu et al. proposed GCL [10] and Chet [11], which employed graph-based methods to establish relationships between diseases and achieved success in health event prediction tasks. However their methods also face certain bottlenecks:

- The clinical observation that if a patient has had disease *A* for a prolonged length of time, the probability of the patient developing disease *B* in the future significantly increases is the primary motivation for employing a graphical approach to establish connections between diseases. Therefore, it is reasonable to draw graphical correlations between diseases with current diagnoses and diseases with high risk. Deep learning models can be used to understand the relationships between high-risk diseases and presently identified diseases, aligning with clinical practice and potentially assisting in disease prediction. However, this clinical experience frequently neglects the potential for future diagnoses of low-risk disorders [14]. While using high-risk diseases to forecast future diseases can improve model performance, ignoring low-risk ones will create a bottleneck in the model. As illustrated in Figure 1, previous methods have focused on the relationship between high-risk diseases and diagnosed diseases; ours looks at not only the relationship between high-risk diseases and diagnosed diseases but also the relationship between low-risk diseases and diagnosed diseases.
- RNNs are frequently used to extract temporal information from a patient's historical diagnostic records after correlations between currently diagnosed diseases and high-risk or low-risk diseases have been shown graphically. This approach aligns with clinical experience, as it involves predicting disease progression based on a patient's prior medical history. However, when dealing with long-term sequences, RNN algorithms encounter the issue of forgetting. Depending on the co-occurrence connections between diseases, it is possible to overlook global information and consider certain diseases as low-risk when they seem to have improved. However, these diseases often have a significant likelihood of recurrence, which will create a bottleneck in the model.

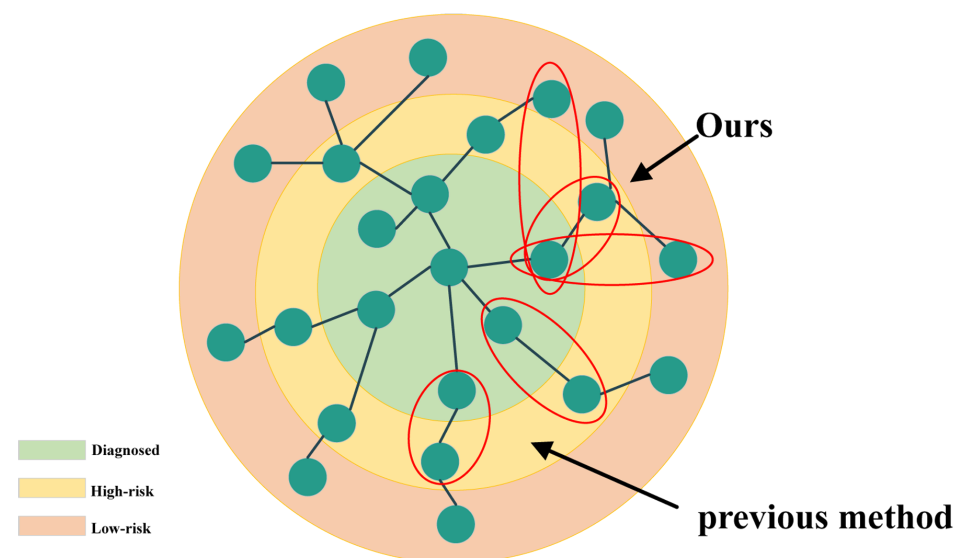


Figure 1. Difference between the previous methods such as GCL [10], Chet [11], and our method DRR. GCL and Chet have only focused on the relationship between high-risk diseases and diagnosed diseases, DRR focuses also on the relationship between low-risk diseases and diagnosed diseases.

We propose two modules to address these problems: the disease relationship reasoning module and the global graph-based feature fusion module. We offer a module for disease relationship reasoning to address the first bottleneck. The aim is to disrupt the co-occurrence associations among specific diseases to increase the number of samples in the training set where low-risk diseases transition to confirmed diseases. This compels the model to focus on the connections between diagnosed diseases and low-risk diseases. Additionally, we introduce a module for global graph-based feature fusion and denoising to tackle the second challenge. This module extracts subgraph-based features for high-risk diseases, presently diagnosed diseases, and their relationships with high-risk diseases at prior time points at various time points. A global subgraph is created by denoising and combining subgraphs from multiple time points. The issue of forgetting global features is solved by the global subgraph-based features, which complement the global characteristics that RNN algorithms overlook.

We conducted experiments with our method on two public datasets. For the health event prediction task, DRR achieved state-of-the-art results, demonstrating improvements of 2.06% and 2.95% over the previous best results on the two datasets, respectively. In summary, the main contributions of this paper are summarized as follows:

- We propose the DRR model, which reconstructs the relationships between diagnosed diseases, high-risk diseases, and low-risk diseases, breaking the model bottleneck caused by existing models' over-reliance on diagnosed and high-risk diseases.
- In our approach, we mitigate the global feature forgetting issue in disease prediction tasks of the GRU method by de-fusing the features of high-risk diseases at different time nodes with the features of diagnosed diseases.
- DRR was tested on two public datasets, Medical Information Mart for Intensive Care III (MIMIC-III) [15] and Medical Information Mart for Intensive Care IV (MIMIC-IV) [16], and achieved state-of-the-art results, thereby confirming the effectiveness of our method.

2. Related Work

2.1. RNN-Type in Health Event Prediction

Health event prediction is a fundamental aspect of the field of medical informatics, encompassing the prediction of individual medical conditions, diseases, or health-related events. Researchers have made significant strides in this field over the years through various data sources and methodologies. EHR is widely used in studies to forecast health events. Researchers use machine learning and deep learning techniques to mine patient demographics, medical histories, diagnostic records, and other EHR data for insightful information. These models are designed to predict diseases, adverse events, or treatment outcomes based on patients' historical data. Dipole [8] incorporated three attention mechanisms into the recurrent neural network (RNN) to model connections between different patient visits. Choi et al. proposed RETAIN [7], an improved RNN model that combines clinical interpretability with high accuracy. DRR constructs relationships between diagnosed diseases and high-risk or low-risk diseases in a patient's single visit using a graph-based method. Subsequently, it utilizes RNN-like methods to extract temporal features from a patient's entire visit history records. While methods exist that leverage CNN features to address the limitations of RNN-based approaches, they fundamentally differ from DRR.

2.2. Graph Method in Health Event Prediction

Graph-based methods have attracted significant attention and demonstrated remarkable success in the field of health event prediction. These approaches represent relationships among medical entities, such as diseases, symptoms, and drugs, as nodes and edges within a graph structure. This enables the extraction of more intricate features and enhances prediction capabilities [17]. In recent times, substantial efforts have been directed towards two primary objectives: establishing relationships within graphs [18,19] and improving the

efficiency of graph neural networks. Notably, GRAM [13] was the pioneering work that successfully integrated graph-based approaches into disease prediction tasks. G-BERT [12] is another notable example, which constructs disease relationships using graphs and subsequently leverages BERT [20] to process these constructed disease relationships. Moreover, CGL [10] delves deep into the exploration of patient–disease interactions and the utilization of medical domain knowledge. In contrast, the Chet method [11] constructs relationships between diseases using graphs and utilizes RNN-like techniques for the extraction of temporal features. This is achieved by incorporating transformation functions that learn patterns in disease progression, ultimately enabling predictions of future disease development. DRR, during the construction of the graph, intentionally disrupts certain co-occurrence relationships to sharpen the model’s attention on connections between diagnosed diseases and low-risk diseases, thus effectively tackling the bottleneck challenges encountered in the aforementioned graph-based methods.

2.3. NLP Method in Health Event Prediction

Recently, researchers have recognized natural language processing (NLP) as a crucial technology for healthcare event prediction. NLP techniques learn from extensive amounts of human text data in a self-supervised manner and have demonstrated significant success when fine-tuned for specific tasks within EHR [5,6]. Nevertheless, such models often demand substantial computational resources. In contrast, DRR, with its modest model parameter size of just 2.3 million, can be deployed at the edge for efficient processing.

3. Method

In this section, we present the details of our proposed method DRR. An overview of DRR is shown in Figure 2.

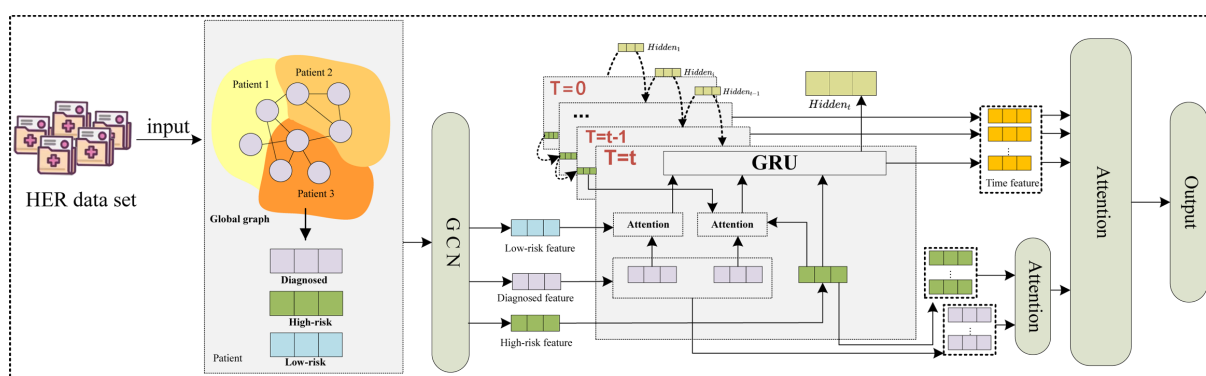


Figure 2. An overview of the proposed DRR model. The model utilizes all HERs to construct a global graph, with nodes symbolizing diseases diagnosed in patients and edges reflecting disease co-occurrence frequencies. In global graph, the node corresponding to the disease diagnosed in patient i at time t is termed the “diagnosed disease node”. Nodes linked to this are “high-risk nodes”, while those connected to high-risk but not to the diagnosed disease node are “low-risk nodes”. The model extracts three types of subgraph encodings for each patient at time T using GCN for feature extraction. An attention mechanism rebuilds relationships between the diagnosed disease at time T low-risk diseases, high-risk diseases at time T , and high-risk diseases at time $T - 1$. These features are then processed through a GRU to extract temporal features. Finally, the model integrates these temporal, high-risk, and diagnosed disease features to predict patient disease diagnosis at time $T + 1$.

3.1. Problem Formulation

EHR contains the medical records of patients for each visit to the hospital. During each medical encounter, patients are diagnosed with one or more diseases, which are represented by a series of medical codes. These confirmed diseases are assigned specific codes predefined by modern medical systems such as ICD-9-CM or ICD-10. For instance, “left heart failure, diabetes mellitus” is assigned the code 428.1, 250 in ICD-9-CM. We use the

set $U = \{u_1, u_2, \dots, u_d\}$ to represent the confirmed diseases, where d is the code number. For each patient $p \in P$, a set of vectors $m^t = \{m_{u_1}, m_{u_2}, \dots, m_{u_d}\}$ is used to represent the diagnosed diseases during the t -th visit, where $m_{u_i}^t \in \{0, 1\}$, and $m_{u_i}^t = 1$ indicates that the patient was diagnosed with disease u_i during the t -th visit, and $t = 1, 2, \dots, T, i \in [1, d]$, where T is the visit number of p . For patient p , all of their diagnostic records can be represented as $E_p = \{m^1, m^2, \dots, m^T\}$. The EHR dataset is defined as $D = \{E_p | p \in P\}$.

Health event prediction involves using an EHR dataset D , given a patient p 's entire history of confirmed medical records E_p , to predict all potential health events m^{t+1} that p may encounter during their next hospital visit.

Common disease prediction involves the identification of prevalent chronic conditions such as palpitations, hypertension, diabetes, etc. By analyzing a patient's historical confirmed medical records E_p , it aims to predict whether the patient will be diagnosed with a specific disease u_i in the future, $m_{u_i}^{T+1} \in \{0, 1\}$.

3.2. Global Graph Definition

We constructed the global graph G using the method proposed by Chet. In G , each node represents a disease u_i from set U . The edges $\langle u_i, u_j \rangle$ in G represent the frequency of diseases u_i and u_j co-occurring. It is important to note that the values of $\langle u_i, u_j \rangle$ and $\langle u_j, u_i \rangle$ are different because, based on clinical experience, the presence of disease u_i may lead to the presence of disease u_j , but the reverse is not necessarily true. We used an adjacency matrix $A \in R^{d \times d}$ to represent the global graph G :

$$A_{ij} = \begin{cases} 0 & \text{if } i = j \text{ or } \frac{f_{ij}}{\sum_{j=1}^d f_{ij}} < \delta, \\ \frac{f_{ij}}{q_i} & \text{otherwise.} \end{cases}, \quad (1)$$

$$\Delta_i = \left\{ c_j \mid \frac{f_{ij}}{\sum_{j=1}^d f_{ij}} \geq \delta \right\}. \quad (2)$$

A_{ij} represents the edges $\langle u_i, u_j \rangle$ in the global graph G . f_{ij} is the frequency of disease u_j appearing in the samples when disease u_i is present. We have set a threshold δ , to filter out diseases that have low-frequency occurrences when disease u_i is present while retaining diseases with high-frequency occurrences Δ_i , q_i represents the sum of frequencies of all other high-frequency occurring diseases when disease u_i is present, $q_i = \sum_{c_j \in \Delta_i} f_{ij}$.

3.3. Disease Relationship Reasoning Module

In this section, we introduce the disease relationship reasoning module in DRR. From the diseases identified in patients at time t , it is possible to deduce high-risk diseases based on clinical experience. In addition, the illnesses that people may be diagnosed with at time $t + 1$ frequently develop from the high-risk illnesses noted at time t . However, the low-risk illnesses discovered at time t may also be the source of the illnesses detected at time $t + 1$. If the majority of the training samples at time $t + 1$ are derived from high-risk diseases identified at time t , the model may not allocate sufficient attention to the relationship between low-risk diseases at time t and the diagnosed diseases at time $t + 1$. To tackle this issue, we employed a masking strategy to conceal certain diseases identified at time t , effectively categorizing some high-risk diseases as low-risk diseases. This method increases the number of instances where low-risk diseases progress into diagnosed diseases, as defined in Equations (3) and (4):

$$mask_i = \begin{cases} 0 & \alpha \\ 1 & 1 - \alpha \end{cases}, 0 \leq \alpha \leq 1, \quad (3)$$

$$\bar{m}^t = m^t \oplus mask. \quad (4)$$

We have set a threshold value α . For a patient u at a certain time t and all confirmed disease sequences m^t for any disease j , there is a probability $1 - \alpha$ that m_j^t transitions from confirmed to unconfirmed, and a probability of α to remain unchanged, $\bar{m}^t$ represents the confirmed sequence that has undergone random masking.

E_i^t represents the disease confirmed in patient i at time t . We need to predict E_i^{t+1} based on E_i^t . If patient j is diagnosed with a disease similar to that of patient i and also develops some diseases that patient i does not have, then we consider these newly developed diseases as high-risk diseases H_i^t that patient i may be diagnosed with at time $t + 1$. It is important to note that some diseases may not have appeared in patient i 's previous visits or in the list of high-risk diseases that may be confirmed, but they appear in the $t + 1$ diagnosis. We refer to such diseases as low-risk diseases L_i^t . This assumption is also based on complex clinical experience.

$$H_i^t = \begin{cases} 1 & m_j^t = 1, m_i^t = 0, A_{ij} \neq 0 \\ 0 & \text{otherwise} \end{cases}, \quad m_j^t, m_i^t \in \bar{m}^t \quad (5)$$

$$L^t = \neg \bar{m}^t \cap \neg H^t \quad (6)$$

We provided the calculation method for high-risk diseases H_i^t for a patient at time t from Equation (5). A patient's high-risk diseases at time t are represented as $H^t = \{H_1^t, H_2^t, \dots, H_d^t\}$. Low-risk $L^t = \{L_1^t, L_2^t, \dots, L_d^t\}$ are determined using Equation (6).

In order to transform the multi-hot column vectors of $\bar{m}_t$, H_t , and L_t into vectors suitable for deep learning model training, we conducted a self-encoding on them in the context of the global graph G . The details are as follows:

$$L_{em}^t = (L^t \odot em), H_{em}^t = (H^t \odot em), \bar{m}_{em}^t = (\bar{m}^t \odot em), \quad (7)$$

$$F_m^t = \partial(W_{\bar{m}}(A \otimes \bar{m}_{em}^t + \bar{m}_{em}^t + H_{em}^t) + b_{\bar{m}}), \quad (8)$$

$$F_H^t = \partial(W_H(A \otimes H_{em}^t + H_{em}^t + \bar{m}_{em}^t) + b_H), \quad (9)$$

$$F_L^t = \partial(W_L(A \otimes L_{em}^t + L_{em}^t) + b_L). \quad (10)$$

where em denotes a set of randomly generated vectors between 0 and 1, and $F_{\{m,H,L\}}^t$ denotes the feature encoding learned through the global graph G at time t . $W_{\{\bar{m},H,L\}}$ and $b_{\{\bar{m},H,L\}}$ are weight and bias matrices, and ∂ refers to the LeakyReLU [21] activation function.

3.4. Global Graph-Based Feature Fusion Module

This section introduces the proposed global graph-based feature extraction module. To improve future disease prediction accuracy, we extract information from the patient's previous diagnostic records. The patient's historical diagnostic features are divided into two categories: local subgraph-based features extracted using our global graph-based feature and fusion module from the patient's subsequent diagnoses, and temporal features extracted using a GRU [22] module from sequential information. To extract global graph-based characteristics representing the patient's disease progression, we aggregate all feature subgraphs from different time points into a unified representation. These global graph attributes will provide essential decision support for identifying potential future illnesses of the patient. The denoising module seeks to combine these feature subgraphs since local feature subgraphs from various time points may contain information that overlaps. This enhances the accuracy of predictions by allowing the global graph-based feature data to complement the temporal features produced by the GRU.

$$F_{GRU}^t, hidden_t = GRU(F_m^t, F_H^t, F_L^t, hidden_{t-1}) \quad (11)$$

$$g_m^t = Covn1d(F_m^t) \quad (12)$$

$$g_H^t = Covn1d(F_H^t) \quad (13)$$

$$g_m^{t-1} = \text{Covn1d}(F_m^{t-1}) \quad (14)$$

$$\delta = \text{softmax}\left(\frac{g_m^t (g_m^{t-1})^T}{\sqrt{a}}\right) g_H^t \quad (15)$$

$$\text{output} = \text{softmax}(\delta W) \quad (16)$$

The confirmed diseases at time t , high-risk diseases, and the high-risk diseases at time $t - 1$ are constructed into a local feature subgraph. The Covn1d is used to denoise the local feature subgraph, followed by the *Atten* module to establish the relationship between high-risk diseases at time $t - 1$, confirmed diseases, and high-risk diseases at time t . The method for establishing the relationship between low-risk diseases and diagnosed diseases is the same as mentioned above. Here, a represents the attention size, g_m^t denotes the features of the confirmed subgraph at time t , g_m^{t-1} signifies the features of the high-risk disease subgraph at time $t - 1$, g_H^t embodies the features of the high-risk disease subgraph at time t , and F_{GAU}^t represents the temporal features extracted using the GRU module at time t . Specifically, g_m^t is utilized as Q, g_m^{t-1} as K, and g_H^t as V to derive the local feature subgraph F_G^t through the application of the *Atten* module. The global graph-based features are denoted as $F_G = \{F_G^1, F_G^2, \dots, F_G^t\}$, and the temporal features as $F_{GAU} = \{F_{GAU}^1, F_{GAU}^2, \dots, F_{GAU}^t\}$. After concatenating these vectors, they are fused using *Atten* to combine the temporal features and global graph-based features. Finally, the fused features are passed through a classification module to obtain the disease diagnosis prediction results at time $T + 1$.

4. Experiments

4.1. Experimental Setups

We use three common tasks to predict health events:

- **Disease prediction.** This task involves predicting all possible diagnosed diseases for a patient at time $T + 1$ based on the patient's previous T instances of confirmed disease records. It is a multi-label classification.
- **Heart failure prediction.** This task involves predicting whether a patient will be diagnosed with heart failure at time $T + 1$ based on the patient's previous T instances of confirmed disease records. It is a binary classification.
- **Common disease prediction.** We have collected data on some common diseases diagnosed in the MIMIC-IV dataset, including hypertension, diabetes, and others. This task involves predicting whether a patient will be diagnosed with these common diseases at time $T + 1$ based on the patient's previous T instances of confirmed disease records. It is a binary classification.

The above three tasks all use *sigmoid* as the activation function for the classifier and employ binary cross-entropy loss as the loss function.

Evaluation metrics. The evaluation metrics for the health events prediction are weighted F_1 score ($w\text{-}F_1$) [23] and top k recall ($R@k$) [24]. $w\text{-}F_1$ is a performance metric for evaluating medical codes, calculated as the weighted sum of F_1 scores. This metric is commonly employed to assess the performance of machine learning models in the medical field. $R@k$ is a metric used to evaluate the performance of predictive models. It assesses the model's performance by calculating the average ratio of correctly predicted medical codes in the top k predictions for each visit or scenario, compared to the total number of correct medical codes. This metric helps evaluate the model's ability to prioritize and predict the most relevant medical codes.

Datasets. We used MIMIC-III [15] and MIMIC-IV [16] to validate the predictive power of DRR. MIMIC-III contains 7493 patients with multiple visits ($T \geq 2$) from 2001 to 2012, while MIMIC-IV includes 85,155 patients with multiple visits from 2008 to 2019. There is a temporal overlap between the two datasets, and we randomly selected 7493 patients from MIMIC-III and 10,000 patients from MIMIC-IV for the years 2013 to 2019.

We divided both the two datasets into training, validation, and test sets based on patient records randomly. In the case of MIMIC-III, this involved 6000/493/1000 patients,

while MIMIC-IV included 8000/1000/1000 patients. We used the last visits as labels and considered the rest as features. The global combination graph G was constructed based on feature visits within the training set.

For the common disease prediction task, we identified common diseases among patients in the MIMIC-IV training set. If a patient is diagnosed with a common disease during their last visit, we assign a label of 1; otherwise, the label is set to 0. The heart failure prediction task followed a similar method.

Baseline methods. To compare DRR with the state-of-the-art models, we selected the following methods as baselines:

- RNN/Attention-based model: RETAIN [7], Dipole [8], Timeline [23], and HiTANet [25].
- CNN-based model: Deepr [9].
- Graph-based model: GRAM [13], G-BERT [12], CGL [10], and Chet [11].

The evaluation metrics for the above-mentioned baseline were based on the results reported in Chet's study.

Parameter settings. In the experiments, we initialized the model's parameters randomly. Hyperparameters and activation functions were tuned on the validation set. For disease prediction, the masking rate α was set to 0.6 for MIMIC-III and 0.5 for MIMIC-IV. The GRU hidden size was set to 256 for MIMIC-III and 200 for MIMIC-IV. Subgraph-based features F_G^t were the same for both datasets, set at 32. Atten size a was set to 32 for both datasets. For heart failure prediction, the masking rate α was set to 0 for both datasets. The GRU hidden size was set to 32 for MIMIC-III and 64 for MIMIC-IV. Subgraph-based features F_G^t were the same for both datasets, set at 32. Atten size a was set to 32 for both datasets. Common disease prediction and heart failure prediction used the same parameter settings, with experiments conducted only on the MIMIC-IV dataset.

We employed the Adam [26] optimizer with 100 iterations and a learning rate of 0.01. Our experiments were carried out on a GeForce RTX 3090 while using Python 3.10 and PyTorch 2.0.

4.2. Comparative Experiments

Disease prediction. Comparison of diagnosis prediction results on the MIMIC-III and MIMIC-IV datasets using w -F1 (%), $R@10$ (%), $R@20$ (%), AUC (%), and $Param$ (M). On average, each patient in the dataset has around 13 diseases per visit, and we chose $k = 10$ and $k = 20$ for our $R@K$ metric to evaluate the performance of disease prediction.

In Tables 1 and 2, we showcase DRR's predictive performance for diseases across both the MIMIC-III and MIMIC-IV datasets. Our model achieves the state-of-the-art performance. Notably, it performs admirably despite having a modest 2.3 million parameters. Furthermore, our model outperforms other models significantly, particularly on the MIMIC-IV dataset. MIMIC-IV is more advanced than MIMIC-III due to its inclusion of a larger volume of historical patient data. This highlights the shortcomings of earlier methodologies in handling extended time series, while our approach successfully augments time-series features by extracting global subgraph-based features.

Heart failure prediction. We showcase the results of our model in the heart failure prediction task in Table 3, using AUC (%) and F_1 (%) as evaluation metrics. Even in low-parameter situations, our method achieves state-of-the-art performance. It is crucial to notice that we set the masking rate to 0 for the heart failure prediction task. This is primarily due to the importance of disease co-occurrence associations in clinical medical diagnosis. We categorize undiscovered illnesses into high-risk and low-risk categories by considering these co-occurrence associations. Although a patient's condition might be caused by a low-risk disease, the majority of confirmed diseases should be caused by high-risk diseases. We use masking to disrupt some of the disease co-occurrence relationships in order to force the model to learn the association between detected diseases and low-risk diseases. It is interesting to note that while the disease prediction task benefits from using masking, the heart failure prediction challenge sees the model performing better without it. This discovery is primarily attributed to the fact that the heart failure prediction task focuses on

determining whether a patient will be diagnosed with heart failure at $T + 1$, making the correlations between diseases' co-occurrences essential. On the other hand, the association between currently diagnosed diseases and low-risk diseases emerges as the key bottleneck for boosting model performance in the disease prediction task, which aims to predict all potential diseases a patient may be diagnosed with at $T + 1$. This outcome also supports the important role played by our disease relationship reasoning module in establishing connections between low-risk diseases and currently identified diseases.

Table 1. Comparison of diagnosis prediction results on MIMIC-III datasets using w -F1 (%), $R@10$ (%), $R@20$ (%), AUC (%), and $Param$ (M). On average, each patient in the dataset has around 13 diseases per visit, and we chose $k = 10$ and $k = 20$ for our $R@K$ metric to evaluate the performance of disease prediction.

Diagnosis Prediction		MIMIC-III		
Models	w -F ₁ (%)	$R@10$ (%)	$R@20$ (%)	Params (M)
RETAIN	20.69	26.13	35.08	2.90
Deeprr	18.87	24.74	33.47	1.16
GRAM	21.52	26.51	35.80	1.59
Dipole	19.35	24.98	34.02	2.18
Timeline	20.46	25.75	34.83	1.23
G-BERT	19.88	25.86	35.31	6.15
HiTANet	21.15	26.02	35.97	3.33
CGL	21.92	26.64	36.72	1.5
Chet	22.63	28.64	37.87	2.12
DRR	24.69	28.31	37.43	2.34

Table 2. Comparison of diagnosis prediction results on MIMIC-IV datasets using w -F1 (%), $R@10$ (%), $R@20$ (%), AUC (%), and $Param$ (M). On average, each patient in the dataset has around 13 diseases per visit, and we chose $k = 10$ and $k = 20$ for our $R@K$ metric to evaluate the performance of disease prediction.

Diagnosis Prediction		MIMIC-IV		
Models	w -F ₁ (%)	$R@10$ (%)	$R@20$ (%)	Params (M)
RETAIN	24.71	28.02	34.46	3.56
Deeprr	24.08	26.29	33.93	1.44
GRAM	23.50	27.29	36.36	1.67
Dipole	23.69	27.38	35.58	2.51
Timeline	25.26	29.00	37.13	1.52
G-BERT	24.49	27.16	35.86	7.53
HiTANet	24.92	27.45	36.37	3.93
CGL	25.41	28.52	37.15	1.83
Chet	26.35	30.28	38.69	2.59
DRR	29.30	30.73	39.65	2.32

Common disease prediction. As shown in Table 4, we selected four common disorders to assess the applicability of our methodology from the MIMIC-IV dataset. To predict these four prevalent diseases, we used the same factors as in the heart failure prediction task. Additionally, we used the state-of-the-art Chet method as a benchmark because it also utilizes parameters from the heart failure prediction task.

Table 3. Heart failure prediction results on MIMIC-III and MIMIC-IV using AUC (%) and F1 (%).

Heart Failure	MIMIC-III			MIMIC-IV		
Models	AUC (%)	F ₁ (%)	Params (M)	AUC (%)	F ₁ (%)	Params (M)
RETAIN	83.21	71.32	1.67	89.02	67.38	1.99
Deeprr	81.36	69.54	0.53	88.43	61.36	0.65
GRAM	83.55	71.78	0.96	89.61	68.94	0.88
Dipole	82.08	70.35	1.41	88.69	68.94	0.88
Timeline	83.34	71.03	0.95	87.53	66.07	0.73
G-BERT	81.50	71.18	3.58	87.26	68.04	3.95
HiTANet	82.77	71.93	2.08	88.10	68.21	3.95
CGL	84.19	71.77	0.55	89.05	69.36	0.60
Chet	86.14	73.08	0.68	90.83	74.14	0.88
DRR	86.33	72.35	0.85	94.30	81.57	1.00

Table 4. Common diseases prediction results on MIMIC-IV using AUC (%) and F1 (%).

Diseases Prediction	Chet		DRR	
Diseases Name	AUC (%)	F ₁ (%)	AUC (%)	F ₁ (%)
Diabetes	83.98	74.55	95.13	87.15
Heart Attack	91.13	61.94	94.11	63.58
Hypertension	84.32	75.22	87.52	77.22
Cardiac Arrhythmia	85.34	32.43	90.03	79.37

4.3. Ablation Study

In Table 5, we performed ablation tests to confirm the efficacy of our suggested disease relationship reasoning module and global graph-based feature fusion module. In the MIMIC-IV dataset's disease prediction task, we compared these experiments. DRR_{rd-} is the model with both modules removed. DRR_{r-} is the model with the disease relationship reasoning module removed, and DRR_{d-} is the model with the global graph-based feature fusion module removed. DRR represents the model without removing either module.

Table 5. The results of ablation experiments are conducted on MIMIC-IV for disease prediction tasks, Using w -F₁ (%) as the evaluation metric.

Model Name	w -F ₁ (%)
DRR _{rd-}	26.35
DRR _{r-}	28.95
DRR _{d-}	28.43
DRR	29.30

4.4. Visualization Analysis

We conducted a visualization analysis in the context of heart failure prediction to validate the efficacy of our model. Prior to applying the classifier, we extracted features and reduced the dimensionality using the t-SNE approach [27]. Subsequently, we labeled the data points and applied color coding, with red indicating diagnosed patients and blue indicating undiagnosed patients. Figure 3 presents the visualization results, where Figure 3a represents the outcomes before training, and Figure 3b represents the results after training. In Figure 3a, the features of both diagnosed and undiagnosed patients seem to cluster together. However, after model training, as shown in Figure 3b, we observe that diagnosed and undiagnosed patients are divided into distinct clusters. This demonstrates the strong classification capability of our model.

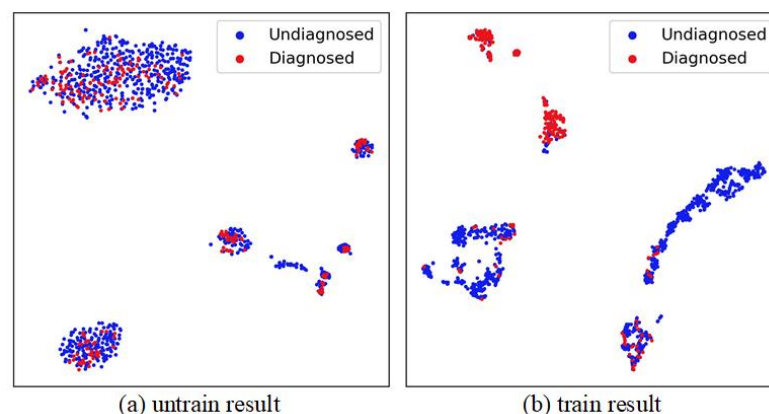


Figure 3. Visualization analysis. Prior to the classifier, features were retrieved, and dimensionality was reduced using the t-SNE approach. The red dots denote the characteristics of patients diagnosed with heart failure, and the blue dots represent those of individuals undiagnosed with the condition. (a) shows the characteristics of patients extracted directly without model training, where features of both diagnosed and undiagnosed individuals are interwoven. In contrast, (b) shows the characteristics post model training, demonstrating a clear demarcation between the features of heart failure patients and those without, thereby emphasizing the model's effectiveness in precise feature differentiation.

5. Conclusions

In this paper, we introduce an approach aiming at addressing the challenges associated with health event prediction. This model incorporates a disease relationship reasoning module and a global graph-based feature fusion module. The disease relationship reasoning module enhances the model's understanding of the relationship between detected diseases and low-risk diseases, thereby overcoming the prediction model's limitations. It achieves this by randomly masking illness co-occurrence connections. The global graph-based feature fusion module complements the global features that are often neglected in RNN-based methods by integrating local subgraph-based features. Experiments conducted on the EHR dataset verify the effectiveness of our method. Our method achieved 2.06% of $w\text{-}F_1$ improvement on MIMIC-III, 2.95% of $w\text{-}F_1$ improvement on MIMIC-IV compared to the current state-of-the-art methods. In the future, we plan to incorporate a wider range of clinical data and explore more efficient and interpretable approaches for healthcare event prediction.

Author Contributions: Conceptualization, Z.L. and Z.D.; methodology, Z.L. and Z.D.; software, Z.L. and Z.D.; validation, Z.L. and Z.D.; Writing—review editing, Z.L., Z.D. and X.L.; visualization, Z.L.; Writing—original draft, Z.D. and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Yunnan Provincial major science and technology project 202102AD080004, 02202AE090019.

Data Availability Statement: The data presented in this study are openly available in reference number.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liang, H.; Tsui, B.Y.; Ni, H.; Valentim, C.C.; Baxter, S.L.; Liu, G.; Cai, W.; Kermany, D.S.; Sun, X.; Chen, J.; et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nat. Med.* **2019**, *25*, 433–438. [[CrossRef](#)] [[PubMed](#)]
2. Henry, J.; Pylypchuk, Y.; Searcy, T.; Patel, V. Adoption of electronic health record systems among US non-federal acute care hospitals: 2008–2015. *ONC Data Brief* **2016**, *35*, 2008–2015.
3. Meystre, S.M.; Savova, G.K.; Kipper-Schuler, K.C.; Hurdle, J.F. Extracting information from textual documents in the electronic health record: A review of recent research. *Yearb. Med. Inform.* **2008**, *17*, 128–144.

4. Yang, J.; Lian, J.W.; Chin, Y.P.H.; Wang, L.; Lian, A.; Murphy, G.F.; Zhou, L. Assessing the prognostic significance of tumor-infiltrating lymphocytes in patients with melanoma using pathologic features identified by natural language processing. *JAMA Netw. Open* **2021**, *4*, e2126337. [[CrossRef](#)]
5. Yang, X.; Chen, A.; PourNejatian, N.; Shin, H.C.; Smith, K.E.; Parisien, C.; Compas, C.; Martin, C.; Costa, A.B.; Flores, M.G.; et al. A large language model for electronic health records. *NPJ Digit. Med.* **2022**, *5*, 194. [[CrossRef](#)] [[PubMed](#)]
6. Patel, R.; Wee, S.N.; Ramaswamy, R.; Thadani, S.; Guruswamy, G.; Garg, R.; Calvanese, N.; Valko, M.; Rush, A.; Rentería, M.; et al. NeuroBlu: A natural language processing (NLP) electronic health record (EHR) data analytic tool to generate real-world evidence in mental healthcare. *Eur. Psychiatry* **2022**, *65*, S99–S100. [[CrossRef](#)]
7. Choi, E.; Bahadori, M.T.; Sun, J.; Kulas, J.; Schuetz, A.; Stewart, W. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. In Proceedings of the 30th Annual Conference on Neural Information Processing Systems, NIPS 2016, Barcelona, Spain, 5–10 December 2016.
8. Ma, F.; Chitta, R.; Zhou, J.; You, Q.; Sun, T.; Gao, J. Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 1903–1911.
9. Wickramasinghe, N. A convolutional net for medical records. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 22–30.
10. Lu, C.; Reddy, C.K.; Chakraborty, P.; Kleinberg, S.; Ning, Y. Collaborative graph learning with auxiliary text for temporal event prediction in healthcare. *arXiv* **2021**, arXiv:2105.07542.
11. Lu, C.; Han, T.; Ning, Y. Context-aware health event prediction via transition functions on dynamic disease graphs. *Proc. AAAI Conf. Artif. Intell.* **2022**, *36*, 4567–4574. [[CrossRef](#)]
12. Shang, J.; Ma, T.; Xiao, C.; Sun, J. Pre-training of graph augmented transformers for medication recommendation. *arXiv* **2019**, arXiv:1906.00346.
13. Choi, E.; Bahadori, M.T.; Song, L.; Stewart, W.F.; Sun, J. GRAM: Graph-based attention model for healthcare representation learning. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 787–795.
14. Schiff, G.D.; Volodarskaya, M.; Ruan, E.; Lim, A.; Wright, A.; Singh, H.; Nieva, H.R. Characteristics of disease-specific and generic diagnostic pitfalls: A qualitative study. *JAMA Netw. Open* **2022**, *5*, e2144531. [[CrossRef](#)] [[PubMed](#)]
15. Johnson, A.E.; Pollard, T.J.; Shen, L.; Lehman, L.W.H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Anthony Celi, L.; Mark, R.G. MIMIC-III, a freely accessible critical care database. *Sci. Data* **2016**, *3*, 160035. [[CrossRef](#)] [[PubMed](#)]
16. Johnson, A.E.W.; Bulgarelli, L.; Shen, L.; Gayles, A.; Shammout, A.; Horng, S.; Pollard, T.J.; Moody, B.; Gow, B.; Lehman, L.-W.H.; et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci. Data* **2023**, *10*, 1. [[CrossRef](#)] [[PubMed](#)]
17. Symeonidis, P.; Kostoulas, T.; Danilatos, V.; Andras, C.; Chairistanidis, S. Mortality Prediction and Safe Drug Recommendation for Critically-ill Patients. In Proceedings of the 2022 IEEE 22nd International Conference on Bioinformatics and Bioengineering (BIBE), Taichung, Taiwan, 7–9 November 2022; pp. 79–84.
18. Li, Y.; Chen, C.; Duan, M.; Zeng, Z.; Li, K. Attention-aware encoder–decoder neural networks for heterogeneous graphs of things. *IEEE Trans. Ind. Inform.* **2020**, *17*, 2890–2898. [[CrossRef](#)]
19. Zou, X.; Li, K.; Chen, C. Multilevel attention based u-shape graph neural network for point clouds learning. *IEEE Trans. Ind. Inform.* **2020**, *18*, 448–456. [[CrossRef](#)]
20. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
21. Xu, B.; Wang, N.; Chen, T.; Li, M. Empirical evaluation of rectified activations in convolutional network. *arXiv* **2015**, arXiv:1505.00853.
22. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078.
23. Bai, T.; Zhang, S.; Egleston, B.L.; Vucetic, S. Interpretable representation learning for healthcare via capturing disease progression through time. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 43–51.
24. Choi, E.; Bahadori, M.T.; Schuetz, A.; Stewart, W.F.; Sun, J. Doctor ai: Predicting clinical events via recurrent neural networks. In Proceedings of the Machine Learning for Healthcare Conference, PMLR, Los Angeles, CA, USA, 19–20 August 2016; pp. 301–318.
25. Luo, J.; Ye, M.; Xiao, C.; Ma, F. Hitanet: Hierarchical time-aware attention networks for risk prediction on electronic health records. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event, CA, USA, 6–10 July 2020; pp. 647–656.
26. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
27. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.