Supplementary Materials for

**miRNA-Disease Association Prediction based on Heterogeneous Graph Transformer with Multi-view similarity and Random Auto-encoder**

Yinbo Liu 1+, Xiaodi Yan 1+, Jun Li 1+, Xinxin Ren 1, Qi Wu 1, Gang-ao Wang1, Yuqing Chen 2 and Xiaolei Zhu 1\*

1 School of Sciences, Anhui Agricultural University, Hefei, Anhui 230036, China

2 School of Mathematics and Statistics, Hefei University of Technology, Hefei 230009, China.

\* Corresponding author: xlzhu\_mdl@hotmail.com

**Supplementary Documents**

**Details of TWMGHT model**

In this paper, we propose a computational method called TWMHGT, which is based on two way encoding modes and Multi-layer Heterogeneous Graph Transforms (MHGT), for miRNA-disease association prediction. The workflow of the proposed method is shown in Figure 1. The first encoding mode is based on the similarity between miRNA and disease from multiple views, while the second encoding mode is based on random auto-encoders. Then, the two sets of encodings are fed into two different MHGT networks to extract high-level embeddings separately. In MHGT, the output encodings from each layer of MHGT are concatenated to form the final encoding. The attention mechanism is then used to fuse the embeddings obtained from the two-way MHGT. Then the matrix multiplication was used to decode the fused embeddings so as to obtain the predicted miRNA-disease sensitivity association matrix. The effectiveness of TWMHGT was evaluated on two different datasets using 5-fold and 10-fold cross-validation. Experimental results demonstrate that our method outperforms existing state-of-the-art approaches.

## Multi-view similarity measures encoding

Previous works have shown that the similarities between miRNAs or between diseases are important information for predicting the associations between miRNAs and diseases. In this study, multi-view similarities of miRNAs or diseases were employed. Based on these multi-view similarities, three different miRNA-miRNA networks and three different disease-disease networks were constructed. For miRNAs, we create three distinct miRNA–miRNA networks by taking miRNA functional similarity (), miRNA sequence similarity () and Gaussian interaction profile kernel similarity matrix () as the adjacency matrices. Similarly, for diseases, we use disease semantic similarity (), target-based disease similarity (), and Gaussian interaction profile kernel similarity matrix () as adjacency matrices to create three different disease–disease networks.

|  |  |
| --- | --- |
| , | (1) |
| . | (2) |

Where denotes functional similarity of miRNAs denotes sequence similarity of miRNAs. Furthermore, and correspond to the similarity matrices obtained by applying the Gaussian interaction profile kernel method to miRNAs and diseases, respectively. Similarly, denotes disease semantic similarity and denotes target-based disease similarity. For detailed calculation process, please refer to the paper of AMHMDA[1].

## Random auto-encoding

We use the Python random module to generate random vectors representing the encoding of miRNAs and diseases. Subsequently, we input each of them into separate auto-encoder models that contain a fully connected layer. The objective of these models is to encode the randomly generated vectors into low-dimensional embeddings that capture the essential features of the original data. The embedding process is defined as follows:

|  |  |
| --- | --- |
| , | (3) |
| . | (4) |

where andare vectors randomly generated to represent the similarity of miRNAs and diseases, respectively. Besides, the process of auto-encoding in a fully connected layer is denoted by the notation . These vectors undergo transformation in the fully connected layer to obtain their final embeddings, which reflect the compressed representations of the original data. As a result of this process, input vectors can be generated for subsequent modules.

## Multi-layer Heterogeneous Graph Transformer

Heterogeneous Graph Transformer (HGT) is a state-of-the-art technique for processing complex graphs containing heterogeneous nodes and edges. The need for such an approach arises from the fact that each node and edge type within the graph is associated with a specific set of features that differ in dimension. The architecture of HGT comprises three primary components: Heterogeneous Mutual Attention, Heterogeneous Message Passing, and Target-Specific Aggregation. Taking miRNA as an example ( Figure S1), the encoding for a single layer of MHGT can be represented as follows.

|  |  |
| --- | --- |
|  | (5) |

where represents the neighbors of the target node (miRNA), and denotes all the edges from source node (disease) to target node . and are encodings representing diseases and miRNAs, respectively. MHGT consists of multiple HGT layers, where represents the layer of an HGT within the MHGT. Multiple attention heads are employed to calculate attention scores for each edge. The specific calculation formula is as follows:

|  |  |
| --- | --- |
| , | (6) |
| , | (7) |
| , | (8) |
| , | (9) |
| . | (10) |

For the -th attention , we project the source node with a linear projection into the -th Key vector :, where denotes the number of attention heads and denotes the dimension size of each attention head's vector. Note that is organized according to the types of the source node , which ensures that each type of node has a distinct linear projection that can effectively capture distribution differences. Similarly, for the target node , we project it into the-th Query vector with a linear projection . Subsequently, the attention vectors from all neighboring nodes are collected and normalized using an edge-softmax function, ensuring that their values sum up to one. The multi-head message calculation is defined as follows:

|  |  |
| --- | --- |
|  | (11) |
|  | (12) |
| . | (13) |

To retrieve the message head , the first step is to project the source node of type- to obtain the message vector using a linear projection : . Next, the model incorporates edge dependency by using a matrix . All message heads are gathered to calculate the message HGT for every pair of nodes.

Finally, using the attention vectors as weights, we take the average of the corresponding messages from the source nodes to obtain the updated vector:

|  |  |
| --- | --- |
| . | (14) |

By using this aggregation process, the model is able to gather information from nearby nodes that have different feature distributions and use it to update the target node. In the final step, HGT performs a linear projection on the updated vector, followed by adding a residual connection.

|  |  |
| --- | --- |
| . | (15) |

In this context, represents a layer with the Softmax activation function. After performing the above steps, we get the HGT layer’s output for the target node . Similarly, the embeddings of diseases can be represented as:

|  |  |
| --- | --- |
| . | (16) |

## MHGT embedding concatenation

The node embeddings of different layers of MHGT are concatenated for a better preservation of information from different levels and enhancing the effectiveness of the model. Note that we employed a fully connected layer as a preprocessing step before passing the encoding matrices to the MHGT for mitigating potential inconsistencies.

|  |  |
| --- | --- |
| , | (17) |
| . | (18) |

Where CH represents concatenated hidden enbedding, and I represents the maximum number of layers in HGT.

## Attention Inner product decoder

The CHs generated from the two-way MHGT are further fused by using an attention mechanism and the fused embedding matrices are decoded through matrix multiplication to obtain the predicted miRNA-disease association matrix as follows:

|  |  |
| --- | --- |
| , | (20) |
| . | (21) |

where and represent the final embeddings of miRNAs and diseases, respectively. And denotes the predicted association matrix between miRNAs and diseases. and respectively represent learnable parameters.

To optimize the model, the cross-entropy loss function is applied to calculate the loss during model training:

|  |  |
| --- | --- |
| . | (22) |

Where and represent the negative and positive link sets of all miRNA-disease pairs, respectively. represents the true labels of the training set partitioned from the dataset, marked as 1 or 0.

## Evaluation metrics

To assess the effectiveness of the proposed model, we are performing 5-fold cross validation (5-CV) and 10-CV experiments on a standard dataset. Negative miRNA-disease sensitivity pairs are randomly generated in equal numbers to match the positive samples. They are subsequently divided into five equally sized subsets. Each subset is being used as the test set once, while the other four subsets are used for training. This process is repeated five times to ensure reliable results. We evaluate the model's performance using seven metrics: area under the receiver operating characteristic curve (AUC), area under the precision-recall curve (AUPR), accuracy, precision, recall, F1-score, and specificity. These metrics are calculated using specific equations (23-27).

|  |  |
| --- | --- |
|  | (23) |
| , | (24) |
| , | (25) |
| , | (26) |
| . | (27) |

**Supplementary Figures**



Figure S1. The single-layer heterogeneous graph transformer structure is applicable for miRNA-disease data.

5-cv超参数

Figure S2. Hyperparameters sensitivity analysis. ((a)-(d) are the results of Data1. (e)-(h) are the results of Data2. (a) and (e) are the different number of hidden channels. (b) and (f) are the different number of attention mechanism heads. (c) and (g) are the different number of MHGT layers. (d) and (h) are the different number of RAEs dimensions.)

1. Ning, Q., et al., *AMHMDA: attention aware multi-view similarity networks and hypergraph learning for miRNA-disease associations identification.* Briefings in Bioinformatics, 2023.