Dear Prof. Guowei Wei,

**RE**: ci-2024-01757a, “Subgraph topology and dynamic graph topology enhanced graph learning and pairwise feature context relationship integration for predicting disease-related miRNAs” Xuan, Ping; Qi, Xiaoying; Chen, Sentao; Gu, Jing; Wang, Xiuju; Cui, Hui; Lu, Jun; Zhang, Tiangang

Thank you very much for the constructive comments and the opportunity to resubmit the revised paper to Journal of Chemical Information and Modeling.

Enclosed is the revised version of the paper and point-by-point responses to the comments and suggestion.

Major revisions include:

* We described the benefit of adaptive fusion of the subgraph enhanced topology and the dynamic graph topology.
* We added the ablation experiments about channel residual strategy and feature residual strategy, the corresponding result analysis was added in the manuscript.
* We added the ablation experiments about channel residual strategy and feature residual strategy, the corresponding result analysis was added in the manuscript.
* We added the experiments to show the results when changing the number of SDGCN encoding layers and the one of FCTransformer encoding layers. We also conducted the experiments to show the impact of channel number within FCTransformer encoding layer on the prediction performance.

We look forward to hearing from you. We greatly appreciate for your effort and help.

Yours sincerely,

Tiangang Zhang on behalf of all the co-authors,

School of Computer Science and Technology, Heilongjiang University, China.

Email: [zhang@hlju.edu.cn](mailto:zhang@hlju.edu.cn)

**Reviewer: 1**

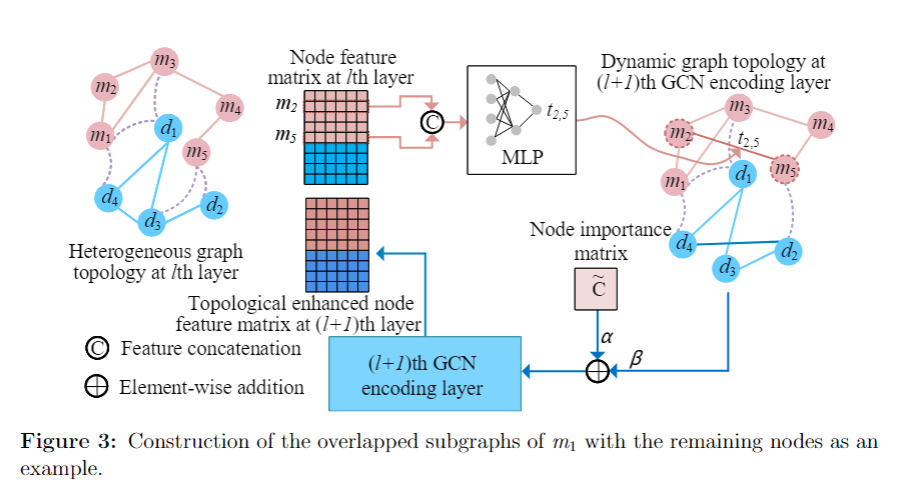
Comments:

Predicting the disease-related miRNAs is helpful for screening the potential candidates for the subsequent biological experiment. In terms of the proposed prediction model, exploiting the neighborhood subgraphs, forming the dynamic graph topology, and encoding the context relationships among the pairwise features are interesting and novel. The extensive comparison experiments, the ablation experiments, and the case studies showed the higher prediction performance of the proposed model, and the effectiveness of its innovations. The suggestions are listed below.

**Response –** We thank the reviewer for the valuable comments. According to the suggestions, we carefully revised the manuscript and provided the following responses. In the revised manuscript, the revisions are highlighted in blue.

1 There is a symbol “” when the dynamic graph topology and the subgraph enhanced topology were fused in Figure 3. Adding its meaning in the figure is necessary.

**Response** –We have added the meaning of “” in Figure 3. The revised image is as follows.



2 and  are the node importance matrix and the dynamic topology matrix, respectively. They are adaptively fused by utilizing the learnable parameters ( and ) in equation 7. Figure 3 also illustrated the corresponding fusion part. Do “+” in the equation and “” in Figure 3 have the same meaning?

**Response** – The symbol “+” in Equation 7 and the symbol “” in Figure 3 have the same meaning. We have replaced the “+” in Equation 7 with “”.

3 What is benefit of adaptively merging the subgraph enhanced topology and the dynamic graph topology? How were the parameters for fusion initialized?

**Response** – Thank you very much for your valuable suggestions. We have added the corresponding description in the fifth paragraph of the “Subgraph and dynamic topologies enhancement for node feature learning” section. The description is also listed as follows.

“ and represent the adjacency matrices of node importance and dynamic topology, respectively. and contribute differently to the topological representation learning of nodes, we designed an adaptive fusion strategy. and are the learnable parameters to balance the contribution of two matrices, resulting in the fused adjacency matrix ,

and are initialized randomly.”

4 The matrices including , and were constructed when the , , were calculated. Please list their dimensions.  
**Response –** The dimensions of​ , and are described in the second paragraph of the “Feature Context Learning Based on Transformer” section. It is also listed as follows.

“In the -th attention head, we conduct 1×1 convolutions on , and then we get , and ,

where represents the dimension of the pairwise feature in the -th layer, and denotes the number of channels per head. , , represent the sets containing 1×1 convolution kernels, respectively.”

5 The feature map was formed by utilizing the activation function, GELU. It is not a very common function, so adding the reference for it is essential.  
**Response** – We have added relevant references to the GELU activation function in the third paragraph of the “Feature Context Learning Based on Transformer” section. It is also listed as follows,

“Here, represents the activation function, is the weight matrix, and is the bias vector.”

**Reference:**

Lee, M. Mathematical analysis and performance evaluation of the gelu activation function in deep learning. Journal of Mathematics 2023, 2023, 4229924.

6 The channel residual strategy and the feature residual one were designed to enhance the pairwise feature context learning. Please add the ablation experiments for both of them to demonstrate their effectiveness.

**Response** – We have added ablation experiments for the channel residual strategy and feature residual strategy. The experimental results and analysis are included in the first and second paragraph of the “Ablation Experiments” section. The description is also listed below in sequence.

“The ablation experiments were conducted to evaluate the effectiveness of inter-node importance evaluation (INIE), dynamic topology embedding (DTE), pairwise feature context learning (PFCL), channel residual strategy (CRS) and feature residual strategy (FRS), as listed in Table 1.”

“After removing CRS, the corresponding AUC and AUPR decreased by 0.4% and 1.6%, respectively. The model without FRS has an AUC and AUPR that are 0.9% and 5.7% lower than the complete model, respectively. This indicated that the supplementation of detailed information from both the channel and feature perspectives is indeed able to enhance the pairwise feature learning.”

Table 1: Results of conducting the ablation experiments on SFPred.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| INIE | DTE | PFCL | CRS | FRS | Average AUC | Average AUPR |
| × | √ | √ | √ | √ | 0.937 | 0.431 |
| √ | × | √ | √ | √ | 0.934 | 0.423 |
| √ | √ | × | √ | √ | 0.908 | 0.296 |
| √ | √ | √ | × | √ | 0.937 | 0.454 |
| √ | √ | √ | √ | × | 0.932 | 0.413 |
| √ | √ | √ | √ | √ | 0.941 | 0.470 |

7 What is the meaning of in equation 20?  
**Response** – In Equation 20, should have been the pairwise vector . We mistakenly wrote as . The correct Equation 20 is listed below:

8 Does the proposed model suffer from the overfitting problem? Which kind of technique was utilized to release its overfitting?  
**Response** – The model experiences overfitting during training, so we adopted the dropout strategy to mitigate it. For the obtained from Equation 19, we applied a fully connected layer with dropout, and the dropout rate is set to 0.5.

**Reviewer: 2**

Comments:  
The authors propose a feature learning model  for predicting disease-related miRNAs by subgraph topology and dynamic graph topology enhanced graph learning and pairwise feature context relationship integration. Although the work is interesting very much, it should be revised from minor comments:

**Response –** We thank the reviewer for the valuable comments. According to the suggestions, we carefully revised the manuscript and provided the following responses. In the revised manuscript, the revisions are highlighted in blue.

1 The authors should further underline main novelties of the proposal model for predicting disease-related miRNAs;

**Response** – Thank you very much for your constructive suggestion. The innovations regarding “inter-node importance evaluation” have been added in the first paragraph of the section “Subgraph and dynamic topologies enhancement for node feature learning”. It is also listed as follows.

“The first-order neighbors of each miRNA (disease) node and the edges among them form the first-order neighbor subgraph. Most of previous methods did not fully leverage the local topology within a subgraph. During the process of graph convolutional learning, the node features change gradually, resulting in a dynamical changing graph topology. We proposed the subgraph and dynamic topology enhanced graph convolutional network (SDGCN) to integrate the topological representations of miRNA and disease nodes.”

The innovations regarding “dynamic topology embedding” have been added in the fourth paragraph of the section “Subgraph and dynamic topologies enhancement for node feature learning”. It is also listed as follows.

“SDGCN consists of multiple graph convolutional layers. As the number of encoding layers increases, the features of miRNA and disease nodes evolve gradually, leading to a dynamic change of the heterogeneous graph topology. Most previous methods did not fully integrate the dynamic topologies during the graph convolution process. We designed a dynamic topology encoding strategy to capture the topological changes of heterogeneous graph.”

We have added the innovation of “pairwise feature context learning” in the first paragraph of the section “Feature context learning based on transformer”. It is also listed as follows.

“A miRNA has the neighboring miRNAs which have similar functions with it and it also has the neighboring diseases which are associated with it. Similarly, a disease also has its neighboring diseases and miRNAs. Thus, the features of a miRNA (disease) include these similarities and associations. As the miRNA with similar functions are usually associated with similar diseases, the pairwise features have contextual relationships. The previous methods did not fully utilize these contextual relationships. We designed a feature context sensitive transformer (FCTransformer) to capture these relationships.”

We have added the innovation related to the "multi-perspective residual strategy" in the third paragraph of the section "Feature context learning based on transformer". It is also listed as follows.

“For the original dual-channel pairwise feature map, the encoding layers of FCTransformer may alter its number of channels and feature dimension. The previous methods did not completely utilize the detailed information from both the channel and feature perspectives. Therefore, it is necessary to design a multi-perspective residual supplementation strategy to complement the information.”

2 The structural parameters of the model should be provided to enhance reproducibility.

**Response** – Based on your suggestions, we have made relevant modifications in the first paragraph of the “Parameter settings” section of the manuscript, as listed below.

“The number of SDGCN encoding layers and that of FCTransformer encoding layers were selected from {1, 2, 3}. The experiments were conducted for all the combinations of their layer numbers. As shown in the supplementary table S1, the model achieved the highest AUC (AUC = 0.941) and AUPR (AUPR = 0.470) when both of their layer numbers were 2. The possible reason is that a single encoding layer failed to completely learn the deep associations between miRNAs and diseases, while three encoding layer may amplify the useless information within the data about miRNAs and diseases. The dimenstions of the output features from SDGCN’s first and second layers were 512 and 128, respectively. The output feature dimensions of FCTransformer’s first and second layers were set to 256 and 64, and the number of heads in each layer was 4. The filter size and number in FCTransformer’s first encoding layer were set to 1×1 and 16, respectively, and its stride was 1. FCTransformer’s second encoding layer contained 16 filters with size=1×1 and stride=1. The channel number of the first layer, , was selected from {128, 64, 32}, and that of the second layer, , was selected from {64, 32, 16}. The model achieved the best performance when and were 64 and 16, respectively (supplementary table S2). After the enhanced pairwise feature representations were obtained, the input and output dimensions of the fully-connected neural network were 1024 and 256, respectively. During the process of estimation of association scores, the dimension of hidden layer of fully-connected neural network was 64. During training, the learning rate was set to 0.001, and weight decay was set to 0.0005.”

Supplementary Table S1. Prediction results for the different layer numbers of SDGCN and FCTransformer.

|  |  |  |  |
| --- | --- | --- | --- |
| SDGCN | FCTransformer | AUC | AUPR |
| 1 | 1 | 0.937 | 0.469 |
| 1 | 2 | 0.936 | 0.446 |
| 1 | 3 | 0.938 | 0.451 |
| 2 | 1 | 0.937 | 0.451 |
| 2 | 2 | 0.941 | 0.470 |
| 2 | 3 | 0.939 | 0.464 |
| 3 | 1 | 0.922 | 0.361 |
| 3 | 2 | 0.928 | 0.391 |
| 3 | 3 | 0.926 | 0.378 |

Supplementary Table S2. Prediction performance when changing the numbers of and

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | AUC | AUPR |
| 128 | 64 | 0.913 | 0.27 |
| 128 | 32 | 0.907 | 0.234 |
| 128 | 16 | 0.916 | 0.312 |
| 64 | 64 | 0.911 | 0.256 |
| 64 | 32 | 0.930 | 0.376 |
| 64 | 16 | 0.941 | 0.470 |
| 32 | 64 | 0.930 | 0.385 |
| 32 | 32 | 0.924 | 0.321 |
| 32 | 16 | 0.936 | 0.412 |

3 Many neural networks have been developed, why did use convolutional transformer learning to improve feature extraction capabilities?

**Response** – Thank you for your comment. First, if miRNA and disease share more similarities and associations with other miRNAs and diseases, is more likely to be associated with . The feature pair corresponding to the -th position in the feature map *F* of the pairwise is more valuable information in relation to and (as shown in Figure 1c). For example, when is 5, the feature pair at this position represents the similarity betweenand , as well as the association between and . If is similar to and is associated with , then is more likely to be associated with . Therefore, we use a 1×1 convolution to focus on learning the features at the -th position. The 1×1 convolution increases the number of feature channels, thereby enriching the features at that position.

Second, the features across multiple channels are contextually related, so we use a transformer-based approach to enhance the connections between them.

We revised the second paragraph of the section “Feature context learning based on transformer” as follows,

“We stack the feature vectors and of and along the channel dimension to obtain the original pairwise feature embedding for (as shown in Figure 1c). The -th column contains a pair of features, and it is worth to be focused on. Assuming is 5, if has similar function to and is associated with, is more likely to be associated with . The -th layer’s input feature map has channels. In the -th attention head, we conduct 1×1 convolutions on , and then we get , and ,

where represents the dimension of the pairwise feature in the -th layer, and denotes the number of channels per head. , , represent the sets containing 1×1 convolution kernels, respectively.”

4 The idea of feature learning has been widely explored in many tasks, and more in-depth research motivation should be analyzed by referring to some references,  
e.g., [https://doi.org/10.1002/jmri.29294;](https://doi.org/10.1002/jmri.29294;https://doi.org/10.1016/j.neunet.2023.05.052;https://doi.org/10.1093/bib/bbac021)[https://doi.org/10.1016/j.neunet.2023.05.052;https://doi.org/10.1093/bib/bbac021](https://doi.org/10.1002/jmri.29294;https://doi.org/10.1016/j.neunet.2023.05.052;https://doi.org/10.1093/bib/bbac021)

**Response** – We have added the relevant description in the fourth paragraph of the “Introduction” section. It is also listed as follows.

“Deep learning is able to learn the deep representative features of objects, and achieve decent performance in image reconstruction for magnetic particle imaging25 and disease-related miRNA prediction. Some approaches utilize neural networks to learn the features of miRNAs and diseases and predict candidate miRNAs through neural-induced matrix completion models26,27. For example, the GCNA-MDA28 prediction model learns the feature representations of miRNA and disease nodes using fully-connected neural networks and autoencoders. However, these models typically focus on individual miRNA similarity graphs and disease similarity graphs. For the miRNA-disease heterogeneous network, several prediction methods have been proposed, which are based on contrastive learning26-32, generative adversarial strategies33,34, graph convolution networks35, stacked autoencoders36, and graph attention networks37-40.”

**Reference:**

Nigam, S.; Gjelaj, E.; Wang, R.; Wei, G.-W.; Wang, P. Machine Learning and Deep Learning Applications in Magnetic Particle Imaging. Journal of Magnetic Resonance

Imaging 2024.

Guo, Y.; Zhou, D.; Ruan, X.; Cao, J. Variational gated autoencoder-based feature extraction model for inferring disease-miRNA associations based on multiview features.Neural Networks 2023, 165, 491–505.

Wang, C.-C.; Li, T.-H.; Huang, L.; Chen, X. Prediction of potential miRNA–disease associations based on stacked autoencoder. Briefings in bioinformatics 2022, 23, bbac021.

5 How did you choose the parameters of the proposed model? The authors could add the necessary parameter analysis.

**Response** – Thank you for your insightful suggestions. The number of layers for SDGCN and FCTransformer are selected from {1, 2, 3}. In FCTransformer, the number of channels for the 1×1 convolution in the first encoding layer is chosen from {128, 64, 32}, and the number of channels for the second layer ranges from {64, 32, 16}. The relevant description has been added to the first paragraph of the “Parameter settings” section. The experimental results for different numbers of layers and channels are listed in supplemental tables S1 and S2. They are also listed as follows.

“The number of SDGCN encoding layers and that of FCTransformer encoding layers were selected from {1, 2, 3}. The experiments were conducted for all the combinations of their layer numbers. As shown in the supplementary Table S1, the model achieved the highest AUC (AUC = 0.941) and AUPR (AUPR = 0.470) when both of their layer numbers were 10. The possible reason is that a single encoding layer failed to completely learn the deep associations between miRNAs and diseases, while three encoding layer may amplify the useless information within the data about miRNAs and diseases.”

“The channel number of the first layer, , was selected from {128, 64, 32}, and that of the second layer, , was selected from {64, 32, 16}. The model achieved the best performance when and are 64 and 16, respectively (supplementary table S2).”

Supplementary Table S1. Prediction results for the different layer numbers of SDGCN and FCTransformer.

|  |  |  |  |
| --- | --- | --- | --- |
| SDGCN | FCTransformer | AUC | AUPR |
| 1 | 1 | 0.937 | 0.469 |
| 1 | 2 | 0.936 | 0.446 |
| 1 | 3 | 0.938 | 0.451 |
| 2 | 1 | 0.937 | 0.451 |
| 2 | 2 | 0.941 | 0.470 |
| 2 | 3 | 0.939 | 0.464 |
| 3 | 1 | 0.922 | 0.361 |
| 3 | 2 | 0.928 | 0.391 |
| 3 | 3 | 0.926 | 0.378 |

Supplementary Table S2. Prediction performance when changing the numbers of and

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | AUC | AUPR |
| 128 | 64 | 0.913 | 0.27 |
| 128 | 32 | 0.907 | 0.234 |
| 128 | 16 | 0.916 | 0.312 |
| 64 | 64 | 0.911 | 0.256 |
| 64 | 32 | 0.930 | 0.376 |
| 64 | 16 | 0.941 | 0.470 |
| 32 | 64 | 0.930 | 0.385 |
| 32 | 32 | 0.924 | 0.321 |
| 32 | 16 | 0.936 | 0.412 |