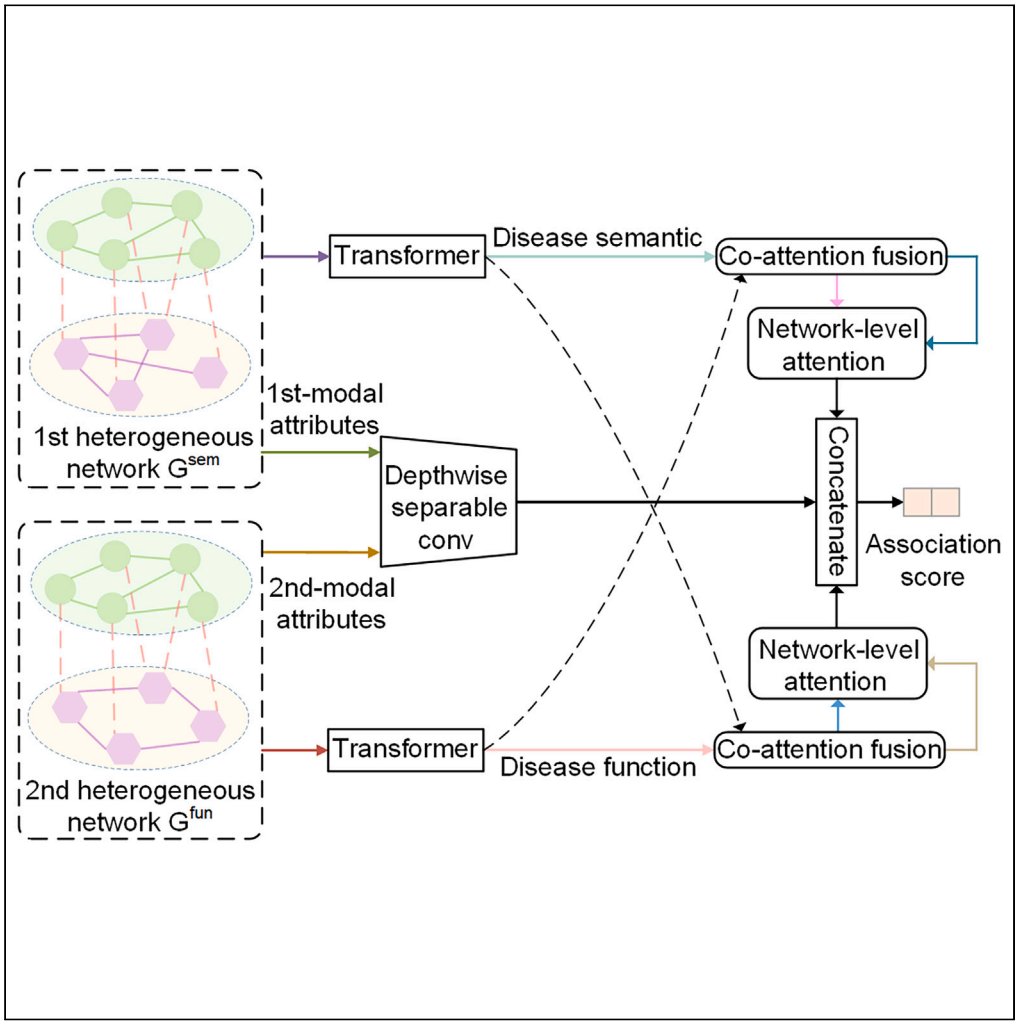
iScience

CellPress
OPEN ACCESS

Article

Complementary feature learning across multiple heterogeneous networks and multimodal attribute learning for predicting disease-related miRNAs



Ping Xuan, Jinshan Xiu, Hui Cui, Xiaowen Zhang, Toshiya Nakaguchi, Tiangang Zhang

zhang@hlju.edu.cn

Highlights

A strategy based on transformer to learn the node feature representations

A co-attention fusion mechanism for encoding the complementary structures and attributes

A network-level attention to discriminate the importance of the features

A depthwise separable convolution strategy to enhance the encoding of each modality data

Xuan et al., iScience 27, 108639
February 16, 2024 © 2023
https://doi.org/10.1016/j.isci.2023.108639

iScience

CellPress
OPEN ACCESS

Article

Complementary feature learning across multiple heterogeneous networks and multimodal attribute learning for predicting disease-related miRNAs

Ping Xuan,[1,2] Jinshan Xiu,[1] Hui Cui,[3] Xiaowen Zhang,[1] Toshiya Nakaguchi,[4] and Tiangang Zhang[1,5,6,*]

SUMMARY

Inferring the latent disease-related miRNAs is helpful for providing a deep insight into observing the disease pathogenesis. We propose a method, CMMDA, to encode and integrate the context relationship among multiple heterogeneous networks, the complementary information across these networks, and the pairwise multimodal attributes. We first established multiple heterogeneous networks according to the diverse disease similarities. The feature representation embedding the context relationship is formulated for each miRNA (disease) node based on transformer. We designed a co-attention fusion mechanism to encode the complementary information among multiple networks. In terms of a pair of miRNA and disease nodes, the pairwise attributes from multiple networks form a multimodal attribute embedding. A module based on depthwise separable convolution is constructed to enhance the encoding of the specific features from each modality. The experimental results and the ablation studies show that CMMDA's superior performance and the effectiveness of its major innovations.

INTRODUCTION

MicroRNAs (miRNAs) are single-stranded non-coding RNAs that include approximately 22 nucleotides that repress the expression of target messenger RNAs (mRNAs).[1–3] Accumulating studies showed that miRNAs are involved in many biological processes, including cell growth, cell proliferation, cellular apoptosis and metabolism.[4–6] Furthermore, miRNAs play a crucial role in the development of various human diseases.[7–9] Calin et al.[10] reported the absent or abnormal expression of the miRNA cluster composed of miR-15 and miR-16 was one factor that causes the chronic lymphocytic leukemia. The researchers demonstrated that the expression level of let-7 in the diabetic patients is obviously lower than that in the normal people.[11] In addition, the concentration of miR-105 was found to be high in the breast tumor cells of the patients with early breast cancer.[12] Therefore, predicting disease-associated miRNA to explore the pathogenesis of diseases is imperative.

Computational predictions of miRNA-disease association can provide biologists with reliable candidate miRNAs for further experimental research. Existing methods are distinguished into three categories. The first is based on the hypothesis that miRNAs with similarity functions are commonly related to similar diseases.[13] The functional similarity between two miRNAs can be calculated through two sets of related diseases,[14] which is then utilized to construct a similarity network of miRNAs. Jiang et al.[15] presented an approach based on such a similarity network and adopted a hypergeometric probability distribution to predict disease-associated miRNA candidates. However, this method only focuses on first-order neighbor information from each node while neglecting multi-order details. Several prediction models are established based on the Random Walk with Restart[16–18] and weighted *k*-neighbor information[19] but are difficult to apply to diseases without known associated miRNAs.

The second category of approaches constructs a heterogeneous network comprised of miRNAs and diseases by introducing additional similarities and associated information related to diseases. You et al.[20] presented a path-based method that exploits a depth-first search algorithm to infer the association propensities between miRNAs and diseases. A matrix decomposition predictive approach was employed[21] that adaptively learns neighbor information for each node to enhance the underlying representation of each node. Additional methods exploit the data of miRNAs and diseases to predict disease-related miRNAs by non-negative matrix factorization,[22,23] random walk,[24–26] and support vector machines.[27] However, these are shallow prediction methods that cannot learn deeper level features of miRNA and disease nodes.

The third category of methods has been proposed that learn deep features of miRNA and disease nodes based on deep learning technology. For example, the prediction model based on graph convolutional network and dual-autoencoder was constructed to learn the

[1]School of Computer Science and Technology, Heilongjiang University, Harbin 150080, China
[2]Department of Computer Science, Shantou University, Shantou 515063, China
[3]Department of Computer Science and Information Technology, La Trobe University, Melbourne, VIC 3083, Australia
[4]Center for Frontier Medical Engineering, Chiba University, Chiba 2638522, Japan
[5]School of Mathematical Science, Heilongjiang University, Harbin 150080, China
[6]Lead contact
*Correspondence: zhang@hlju.edu.cn
https://doi.org/10.1016/j.isci.2023.108639

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

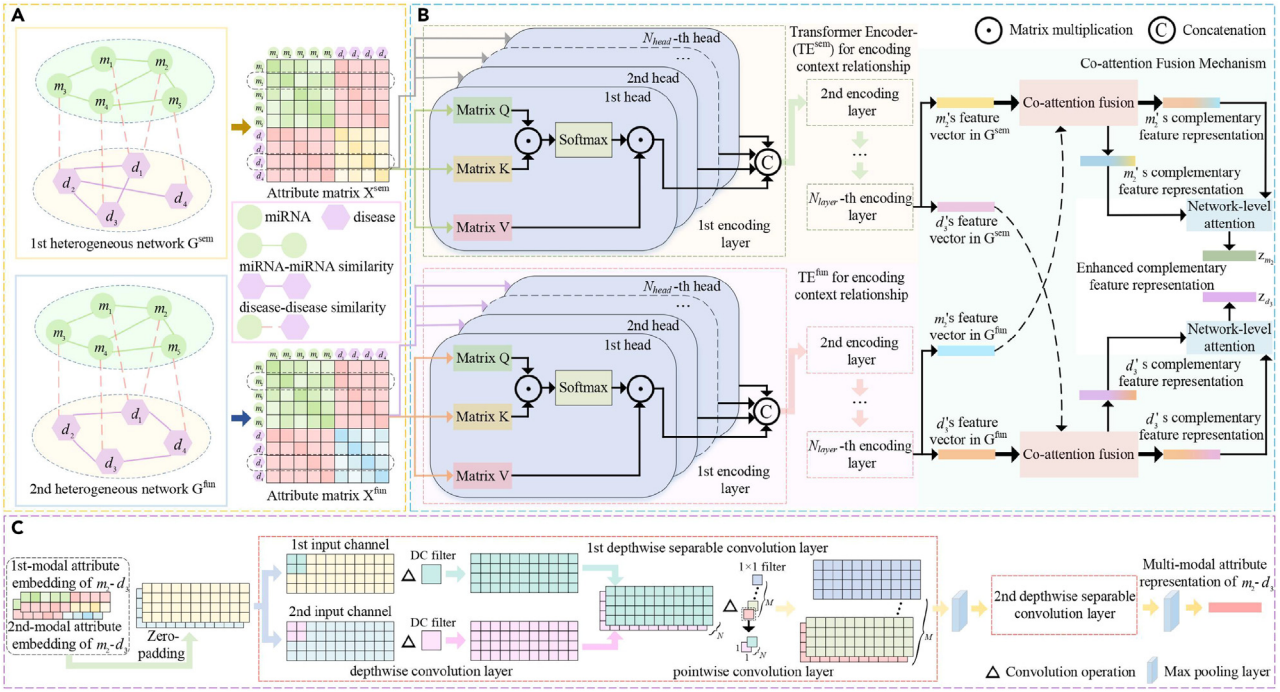


Figure 1. Framework of the proposed CMMDA model

(A) Construct two miRNA-disease heterogeneous networks according to two kinds of disease similarities.

(B) Learn the node feature representations embedding the context relationship and complementary features.

(C) Encode the pairwise multi-modal attributes.

topological features of the miRNA and disease nodes.[28] Other approaches leverage deep belief networks,[29] graph attention networks,[30–32] generative adversarial networks[33] and autoencoders[34–37] that only target a single miRNA-disease heterogeneous network. However, multiple similarities can be calculated between disease nodes, so we can construct multiple miRNA-disease heterogeneous networks. Xie et al.[38] established a predictive method based on graph convolutional networks to learn topological information about miRNAs and diseases from multiple heterogeneous networks but does not consider the complementarity of topological information of multiple networks. Moreover, specific modal attribute embeddings exist for each pair of miRNA and disease nodes in a heterogeneous network, and the attribute embeddings from multiple networks form multimodal attribute embeddings of each node pair. Previous methods do not integrate such multimodal information of node pairs.

We propose a prediction method called CMMDA to learn the node features from across multiple miRNA-disease heterogeneous networks and encode a multimodal attribute from the pairwise node level (Figure 1). The contributions of our approach include the following.

- Two heterogeneous networks composed of the miRNA and disease nodes were established according to multiple kinds of disease similarities. There are context relationships between the attributes of each target miRNA (disease) node and the ones of each of the other miRNA (disease) node within each network. A separate learning module was constructed for each network to embed these context relationships and encode the specific attributes of each miRNA (disease) node in the network.
- Two miRNA-disease networks reflect the similarities and associations among the miRNA and disease nodes from different perspectives. Therefore, complementing the attribute information from one heterogeneous network to the attribute learning process for another network is necessary. We designed a co-attention fusion mechanism which propagates the attribute information from one network to another one to integrate the complementary node attributes (Figure 2).
- The multiple modality attributes of a pair of miRNA and disease nodes derived from multiple miRNA-disease networks. We constructed a module based on depthwise separable convolution networks to learn the modality-sensitive information and then deeply fused the information from multiple modalities.
- As the node attributes of a network and the complementary attributes from another heterogeneous network have discriminative contributions for the miRNA-disease association prediction, we designed an attention at network level to obtain the informative attributes. The improved prediction performance of our method was shown by comparison with several state-of-the-art approaches and the ability of retrieving the latent candidate miRNA-disease associations was confirmed by case studies over 3 diseases. The ablation study results showed the effectiveness of node context feature learning, information interaction across multiple networks, and pairwise multi-modal attribute learning for the improved association prediction performance.

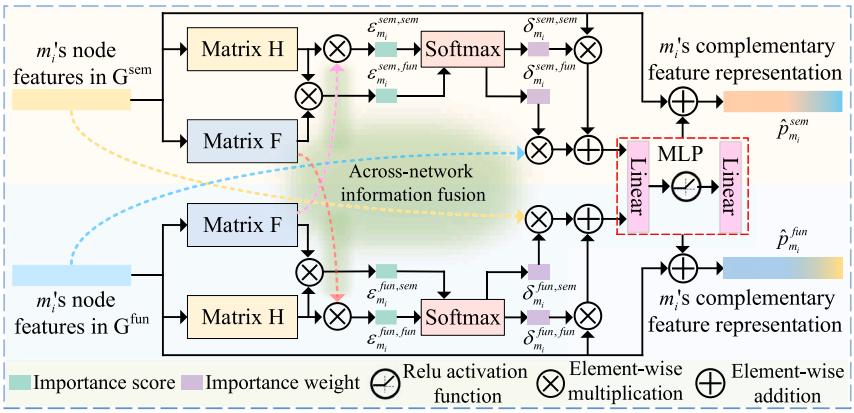


Figure 2. Illustration of the co-attention fusion mechanism which encodes the node complementary information across multiple networks

RESULTS AND DISCUSSION

Evaluation metrics

Five-fold cross-validation evaluated the prediction performance of CMMDA and comparative methods. Positive samples included 12,446 known miRNA-disease associations, with 491,677 unknown associations as negative samples. We randomly divided the positive samples into five sets. For each round, four sets of positive samples and randomly selected negative samples of equal size were treated as training set, and the remaining set of positive samples and all the remaining negative samples were employed to test. The functional similarity between diseases was recalculated by utilizing the known miRNA-disease associations for each fold in the cross-validation. Because a significant imbalance exists between the negative and positive samples, we adopt the area under the receiver operating characteristic (ROC) curve (AUC)[39] and the area under the precision-recall curve (AUPR)[40] for the performance measure of the prediction model. First, we calculated the AUC and AUPR for each disease during each round, then calculated the average value of the 591 diseases during cross-validations, all of which were averaged for a final value. Because biologists are often concerned with the top-ranked prediction results, we calculated the recall rate of the top $k \in [30, 240]$ candidates.

Parameter settings

Our learning model is implemented using the PyTorch framework on an NVIDIA GeForce GTX 2080Ti graphic card with 16 GB memory. For the node feature learning module based on the transformer, both encoders include 4 encoding layers, each containing 8 attention heads with output feature dimensions set to 1,444. The multimodal attribute encoding module comprises two depthwise convolution-pointwise convolution layers. The number of input channels of depthwise convolution in the first and second layers is 2 and 32, respectively, with convolution filter sizes of 3 × 3 and 2 × 3, respectively, and zero padding and step size values of 1 for each. The pooling window size in the two max pooling layers is 1 × 2. The value of γ is set to 0.4 in our experiment.

Ablation studies

We performed ablation studies to evaluate the contributions of the node context relationship encoding (NCRE), co-attention (CA), information fusion across multiple networks (IFAMN), network-level attention (NA) and multimodal attribute encoding (MMAE) on the CMMDA performance. As shown in Table 1, the model containing NCRE, CA, IFAMN, NA and MMAE achieves the best performance. Without NCRE, the AUC and AUPR results were reduced by 0.6% and 2.1%, respectively, compared to the full prediction model. When CA was removed from the training process, AUC and AUPR reduced by 0.9% and 3.2%, respectively. In terms of CA without IFAMN, its AUC and AUPR are reduced by 0.7% and 2.6%, respectively. For the model without NA, its AUC and AUPR decreased by 0.5% and 1.7% when it was compared with CMMDA. Compared to the model without MMAE, the entire model improved by 1.2% in AUC and 4.3% in AUPR. The MMAE provided the most significant contribution to improving the predictive performance of CMMDA because the introduction of the multimodal information more comprehensively reflects the features of the miRNA-disease node pairs, and this module could better learn the specific information for each input channel. In addition, we designed another instance which replaced the MMAE with a traditional convolution (TC) to learn the pairwise attributes. The instance got 0.929 of AUC and 0.407 of AUPR. AUC and AUPR of the instance with MMAE are 1.1% and 2.8% higher than the one with TC, respectively. It indicated that the pairwise attribute learning based on depthwise separable convolution plays important role for the improved prediction performance.

Comparison with other methods

Eleven state-of-the-art methods were compared with CMMDA, including GMDA,[33] PBMDA,[21] DMPred,[22] GSTRW,[26] Liu's method,[25] CFSAEMDA,[35] GCNA-MDA,[28] DAEMKL,[36] DFELMDA,[37] DBNMDA,[29] and AEMDA.[34] As presented in Figure 3, CMMDA acquired the best

Table 1. Results of the ablation studies

NCRE	CA	IFAMN	NA	MMAE	TC	Average AUC	Average AUPR
×	✓	✓	✓	✓	×	0.934	0.414
✓	×	×	✓	✓	×	0.931	0.403
✓	✓	×	✓	✓	×	0.933	0.409
✓	✓	✓	×	✓	×	0.935	0.418
✓	✓	✓	✓	×	×	0.928	0.392
✓	✓	✓	✓	×	✓	0.929	0.407
✓	✓	✓	✓	✓	×	0.940	0.435

average AUC of 0.940, which is 0.9% better than DFELMDA, 1.1% better than GMDA, 7.9% better than PBMDA, 4.8% better than DMPred, 2.1% better than CFSAEMDA, 2.4% better than GCNA-MDA, 4.4% better than DAEMKL, 12.7% better than GSTRW, 4.6% better than Liu's method, 3.1% better than DBNMDA, and 2.2% better than AEMDA. Considering the average AUPR, CMMDA obtained the best AUPR of 0.435, which is 5.9%, 17.5%, 34.4%, 34.8%, 18.8%, 25.5%, 27.1%, 39.1%, 33.4%, 24.5%, and 21.1% higher than DFELMDA, GMDA, PBMDA, DMPred, CFSAEMDA, GCNA-MDA, DAEMKL, GSTRW, Liu's method, DBNMDA, and AEMDA, respectively.

Our CMMDA model provides state-of-the-art performance in terms of AUC and AUPR. DFELMDA designed an ensemble learning model based on deep forest and GMDA constructed a model based on generative adversarial network. They achieved the second-best and the third-best performances, respectively. Both CFSAEMDA and AEMDA designed their models based on autoencoders, and they obtained similar AUC values. Moreover, the former got 2.3% higher AUPR than the latter. DAEMKL fused multiple kinds of miRNA similarity information and disease similarity information based on multiple kernel learning. Its prediction performance is not as good as that of GCNA-MDA. The possible reason is GCNA-MDA captured the topological information of the miRNA and disease nodes by graph neural network. DBNMDA outperformed DMPred and PBMDA as it applied pre-train restricted Boltzmann machine to extract the information of the miRNA-disease node pairs. Liu's method and GSTRW leverage a model with a random walk and are both shallow predictive models that cannot deeply explore the association between miRNAs and diseases. CMMDA achieves the best performance because it fully interacts with the information from the two networks and deeply learns the multimodal information of the miRNA-disease node pairs.

The recall rates of the top *k* miRNA candidates are given in Figure 4, where a higher recall indicates that more disease-associated miRNAs are correctly identified. When *k* increases from 30 to 240 through increments of 30, CMMDA is superior to all other methods at various *k* thresholds. When *k* is 30, 60, and 90, the recall of CMMDA is 58.3%, 81.4%, and 95.8%, respectively. The second-best method of DFELMDA has recall rates of 54.3%, 78.3%, and 93.8% at the top 30, 60, and 90 thresholds. GMDA ranked third with recall rates of 51.4%, 75.8%, and 91.2%, respectively. The recall of CFSAEMDA and AEMDA are similar through these thresholds with rates of 48.3%, 73.5%, and 89.4% and 46.2%, 72.8%, and 88.9%, respectively. GCNA-MDA consistently outperformed DBNMDA with recall rates of 46.7%, 72.0%, and 88.4%, compared to 45.4%, 69.7%, and 82.5% for DBNMDA, respectively. DAEMKL was slightly higher than DMPred, and the corresponding recall rates of DAEMKL were 45.9%, 70.4%, and 82.1%, whereas the recall rates of DMPred are 45.1%, 69.2%, and 81.9%, respectively. The

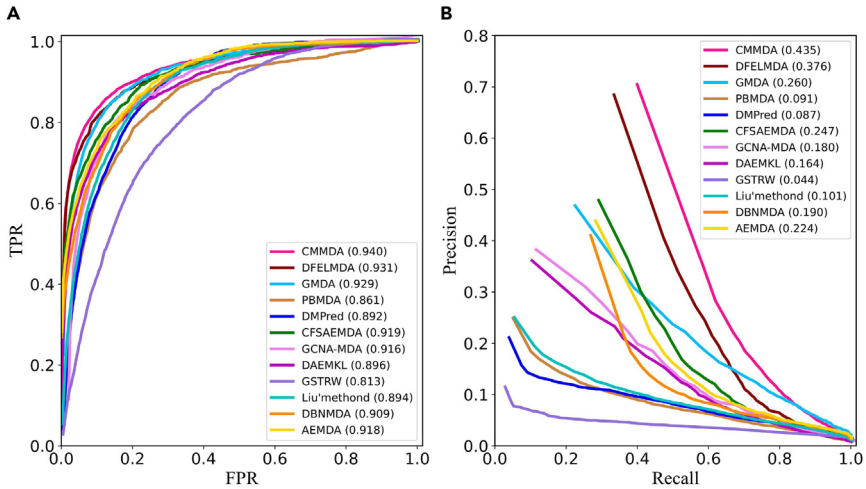


Figure 3. ROC and PR curves of our method and the compared methods
(A) ROC curves; (B) PR curves.

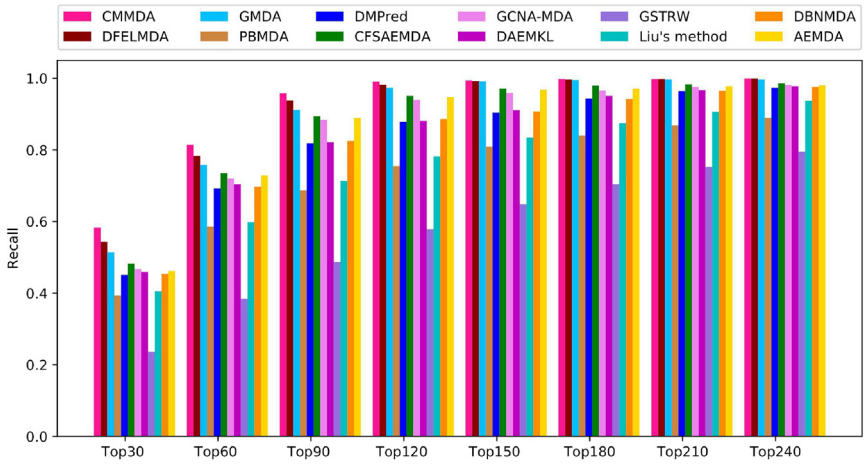


Figure 4. Average recall rates over all the diseases at different k values

recall rates of the Liu's method at 40.5%, 59.8%, and 71.3% are inferior to those of PBMDA at 39.3%, 58.6%, and 68.7%. The lowest recall values are seen with GSTRW as 23.6%, 38.4%, and 48.7%.

Case studies: Esophageal neoplasm, thyroid neoplasm, and kidney neoplasm

To evaluate the capability of CMMDA to identify real-world potential candidates, we conducted case studies with the esophageal neoplasm, thyroid neoplasm, and kidney neoplasm diseases. Table 2 details the top 50 candidate miRNAs for esophageal neoplasm, and Tables S1 and S2 list the top 50 candidates for thyroid neoplasm and kidney neoplasm, respectively. Validation was performed based on three important miRNA-disease association databases, including dbDEMC[41] and miRCancer.[42]

The dbDEMC database has differentially expressed miRNAs in human cancers, including 2,224 miRNAs and 36 cancer types. As seen in Table 2, among the miRNA candidates associated with esophageal neoplasm, 47 candidates appeared in dbDEMC, suggesting that these miRNAs are up-regulated or down-regulated in esophageal neoplasm. Xie et al.[42] applied text mining techniques to extract the empirically verified miRNA-cancer associations from published papers, which were recorded in the miRCancer database. From our results, 23 candidates were contained in the miRCancer sources, suggesting that these predicted candidate miRNAs are associated with esophageal neoplasm. Two miRNA candidates are not supported by corresponding data and are labeled "unconfirmed".

Of the miRNA candidates predicted to be associated with thyroid neoplasm listed in Table S1, 46 are contained in dbDEMC and 12 in miRCancer, again suggesting a strong capability of CMMDA in identifying candidate miRNAs for thyroid neoplasms. The miRNA candidates predicted for kidney neoplasms, as listed in Table S2, had 49 listed in dbDEMC and 9 in miRCancer. There are 4 candidates in the Table S1 and 1 candidate in Table S2 with label "unconfirmed". The label means there are no evidences to be found to confirm the associations among these candidate miRNAs and the corresponding diseases.

Prediction of miRNA-disease associations

Following the evaluation of CMMDA by cross-validation and case studies, as described above, the model was applied to predict miRNA-disease associations. All known miRNA-disease associations and same amount of randomly selected unknown associations were used to train CMMDA model. The top 50 miRNA candidates for each disease predicted by CMMDA with this study are presented in Table S3.

Limitations of the study

Recently, the various information about the miRNAs and diseases appeared, such as the interactions among miRNAs and lncRNAs. Therefore, deep integration of these multi-source data and information and the multimodal data is our future work.

Conclusions

A disease-related miRNA prediction method was proposed to deeply integrate the context relationships among the node attributes within each heterogeneous network and the complementary node features across multiple networks. The constructed two miRNA-disease heterogeneous networks were helpful for the subsequent node context relationship encoding. The designed co-attention fusion strategy was able to exchange the complementary node features across multiple heterogeneous networks. The network-level attention was constructed to assign greater weight to the more important node complementary features from multiple networks. The module based on depthwise separable convolution networks was developed to deeply fuse the multi-modal attributes for each pair of miRNA and disease nodes. The 5-fold cross validation results showed that CMMDA's AUC, AUPR, and its recall rates for the top-ranked candidates were consistently superior to the

Table 2. Top 50 miRNA candidates related to esophageal neoplasm

Rank	miRNA name	Evidence	Rank	miRNA name	Evidence
1	hsa-mir-34a	dbDEMC, miRCancer	26	hsa-mir-141	dbDEMC, miRCancer
2	hsa-mir-135b	dbDEMC	27	hsa-mir-100	dbDEMC, miRCancer
3	hsa-mir-21	dbDEMC, miRCancer	28	hsa-mir-193b	dbDEMC
4	hsa-mir-200c	dbDEMC	29	hsa-mir-26a-1	dbDEMC
5	hsa-let-7a-2	dbDEMC	30	hsa-mir-183	dbDEMC, miRCancer
6	hsa-mir-200a	dbDEMC, miRCancer	31	hsa-mir-148a	dbDEMC
7	hsa-let-7a-3	unconfirmed	32	hsa-mir-302a	dbDEMC
8	hsa-mir-31	dbDEMC, miRCancer	33	hsa-mir-373	dbDEMC
9	hsa-mir-196a-2	unconfirmed	34	hsa-mir-451a	dbDEMC
10	hsa-let-7g	dbDEMC	35	hsa-mir-26a-2	dbDEMC
11	hsa-let-7b	dbDEMC	36	hsa-mir-197	dbDEMC, miRCancer
12	hsa-mir-22	dbDEMC, miRCancer	37	hsa-mir-150	dbDEMC, miRCancer
13	hsa-mir-195	dbDEMC, miRCancer	38	hsa-mir-122	dbDEMC
14	hsa-mir-145	dbDEMC, miRCancer	39	hsa-mir-301a	dbDEMC
15	hsa-let-7a-1	miRCancer	40	hsa-mir-204	dbDEMC, miRCancer
16	hsa-mir-99a	dbDEMC	41	hsa-mir-342	dbDEMC
17	hsa-mir-126	dbDEMC, miRCancer	42	hsa-mir-429	dbDEMC
18	hsa-mir-27b	dbDEMC	43	hsa-mir-92a-2	dbDEMC
19	hsa-mir-143	dbDEMC, miRCancer	44	hsa-mir-34b	dbDEMC
20	hsa-mir-34c	dbDEMC	45	hsa-mir-15a	dbDEMC, miRCancer
21	hsa-let-7i	dbDEMC	46	hsa-mir-19a	dbDEMC
22	hsa-mir-223	dbDEMC, miRCancer	47	hsa-mir-625	dbDEMC
23	hsa-mir-99b	dbDEMC, miRCancer	48	hsa-mir-210	dbDEMC, miRCancer
24	hsa-mir-425	dbDEMC	49	hsa-mir-146b	dbDEMC, miRCancer
25	hsa-mir-205	dbDEMC, miRCancer	50	hsa-mir-96	dbDEMC, miRCancer

compared advanced methods. In addition, case studies on three diseases further demonstrated CMMDA's ability in discovering the latent disease-related candidate miRNAs.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Dataset
  - Multi-source data representations for miRNAs and diseases
  - Capturing context relationships and learning node complementary feature representations
  - Multi-modal attribute embedding of miRNA-disease node pairs
  - Multi-modal attribute encoding by depthwise separable convolution
  - Final integration and optimization
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108639.

ACKNOWLEDGMENTS

The work was supported by the Natural Science Foundation of China (62172143, 61972135, 62372282); STU Scientific Research Initiation Grant (NTF22032); and the Natural Science Foundation of Heilongjiang Province (LH2023F044).

All authors contributed to the article and approved the submitted version.

AUTHOR CONTRIBUTIONS

P.X. designed the method and participated in manuscript writing. J.X. designed the experiments and edited the manuscript. H.C. participated in method design and manuscript writing. X.Z. participated in method design. T.N. participated in experiment design. T.Z. participated in method design and manuscript writing.

DECLARATION OF INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received: July 18, 2023
Revised: November 2, 2023
Accepted: December 1, 2023
Published: December 5, 2023

REFERENCES

1. Ambros, V. (2004). The functions of animal microRNAs. Nature *431*, 350–355.
2. Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281–297.
3. Van Meter, E.N., Onyango, J.A., and Teske, K.A. (2020). A review of currently identified small molecule modulators of microRNA function. Eur. J. Med. Chem. *188*, 112008.
4. Ebert, M.S., and Sharp, P.A. (2012). Roles for microRNAs in conferring robustness to biological processes. Cell *149*, 515–524.
5. Le, T.D., Liu, L., Zhang, J., Liu, B., and Li, J. (2015). From miRNA regulation to miRNA–TF co-regulation: computational approaches and challenges. Brief. Bioinform. *16*, 475–496.
6. Yi, H.C., You, Z.H., Huang, D.S., Li, X., Jiang, T.H., and Li, L.P. (2018). A deep learning framework for robust and accurate prediction of ncRNA-protein interactions using evolutionary information. Mol. Ther. Nucleic Acids *11*, 337–344.
7. Chen, X., Xie, D., Zhao, Q., and You, Z.H. (2019). MicroRNAs and complex diseases: from experimental results to computational models. Brief. Bioinform. *20*, 515–539.
8. Rupaimoole, R., and Slack, F.J. (2017). MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. Nat. Rev. Drug Discov. *16*, 203–222.
9. Metzinger-Le Meuth, V., and Metzinger, L. (2019). miR-223 and other miRNA's evaluation in chronic kidney disease: innovative biomarkers and therapeutic tools. Noncoding. RNA Res. *4*, 30–35.
10. Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K., et al. (2002). Nonlinear partial differential equations and applications: Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc. Natl. Acad. Sci. USA *99*, 15524–15529.
11. Brennan, E., Wang, B., McClelland, A., Mohan, M., Marai, M., Beuscart, O., Derouiche, S., Gray, S., Pickering, R., Tikellis, C., et al. (2017). Protective effect of let-7 miRNA family in regulating inflammation in diabetes-associated atherosclerosis. Diabetes *66*, 2266–2277.
12. Yan, W., Wu, X., Zhou, W., Fong, M.Y., Cao, M., Liu, J., Liu, X., Chen, C.H., Fadare, O., Pizzo, D.P., et al. (2018). Cancer-cell-secreted exosomal miR-105 promotes tumour growth through the MYC-dependent metabolic reprogramming of stromal cells. Nat. Cell Biol. *20*, 597–609.
13. Zhao, Y., Chen, X., and Yin, J. (2019). Adaptive boosting-based computational model for predicting potential miRNA-disease associations. Bioinform *35*, 4730–4738.
14. Wang, D., Wang, J., Lu, M., Song, F., and Cui, Q. (2010). Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. Bioinform *26*, 1644–1650.
15. Jiang, Q., Hao, Y., Wang, G., Juan, L., Zhang, T., Teng, M., Liu, Y., and Wang, Y. (2010). Prioritization of disease microRNAs through a human phenome-microRNAome network. BMC Syst. Biol. *4*, S2–S9.
16. Chen, X., Liu, M.X., and Yan, G.Y. (2012). RWRMDA: predicting novel human microRNA–disease associations. Mol. Biosyst. *8*, 2792–2798.
17. Chen, X., and Yan, G.Y. (2014). Semi-supervised learning for potential human microRNA-disease associations inference. Sci. Rep. *4*, 5501.
18. Li, A., Deng, Y., Tan, Y., and Chen, M. (2021). A novel miRNA-disease association prediction model using dual random walk with restart and space projection federated method. PLoS One *16*, e0252971.
19. Xuan, P., Han, K., Guo, M., Guo, Y., Li, J., Ding, J., Liu, Y., Dai, Q., Li, J., Teng, Z., and Huang, Y. (2013). Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors. PLoS One *8*, e70204.
20. You, Z.H., Huang, Z.A., Zhu, Z., Yan, G.Y., Li, Z.W., Wen, Z., and Chen, X. (2017). PBMDA: A novel and effective path-based computational model for miRNA-disease association prediction. PLoS Comput. Biol. *13*, e1005455.
21. Zhang, Z.C., Zhang, X.F., Wu, M., Ou-Yang, L., Zhao, X.M., and Li, X.L. (2020). A graph regularized generalized matrix factorization model for predicting links in biomedical bipartite networks. Bioinform *36*, 3474–3481.
22. Zhong, Y., Xuan, P., Wang, X., Zhang, T., Li, J., Liu, Y., and Zhang, W. (2018). A non-negative matrix factorization based method for predicting disease-associated miRNAs in miRNA-disease bilayer network. Bioinform *34*, 267–277.
23. Peng, L., Yang, C., Huang, L., Chen, X., Fu, X., and Liu, W. (2022). RNMFLP: predicting circRNA–disease associations based on robust nonnegative matrix factorization and label propagation. Brief. Bioinform. *23*, bbac155.
24. Liu, Y., Zeng, X., He, Z., and Zou, Q. (2017). Inferring microRNA-disease associations by random walk on a heterogeneous network with multiple data sources. IEEE/ACM Trans. Comput. Biol. Bioinform. *14*, 905–915.
25. Chen, M., Liao, B., and Li, Z. (2018). Global similarity method based on a two-tier random walk for the prediction of microRNA–disease association. Sci. Rep. *8*, 6481.
26. Xie, X., Wang, Y., He, K., and Sheng, N. (2023). Predicting miRNA-disease associations based on PPMI and attention network. BMC Bioinform *24*, 113.
27. Chen, X., Huang, L., Xie, D., and Zhao, Q. (2018). EGBMMDA: extreme gradient boosting machine for MiRNA-disease association prediction. Cell Death Dis. *9*, 3.
28. Liao, Q., Ye, Y., Li, Z., Chen, H., and Zhuo, L. (2023). Prediction of miRNA-disease associations in microbes based on graph convolutional networks and autoencoders. Front. Microbiol. *14*, 1170559.
29. Chen, X., Li, T.H., Zhao, Y., Wang, C.C., and Zhu, C.C. (2021). Deep-belief network for predicting potential miRNA-disease associations. Brief. Bioinform. *22*, bbaa186.
30. Li, G., Fang, T., Zhang, Y., Liang, C., Xiao, Q., and Luo, J. (2022). Predicting miRNA-disease associations based on graph attention

network with multi-source information. BMC Bioinform *23*, 244.

31. Wang, W., and Chen, H. (2022). Predicting miRNA-disease associations based on graph attention networks and dual Laplacian regularized least squares. Brief. Bioinform. *23*, bbac292.
32. Zhao, H., Li, Z., You, Z.H., Nie, R., and Zhong, T. (2023). Predicting Mirna-Disease Associations Based on Neighbor Selection Graph Attention Networks. IEEE/ACM Trans. Comput. Biol. Bioinform. *20*, 1298–1307.
33. Xuan, P., Wang, D., Cui, H., Zhang, T., and Nakaguchi, T. (2022). Integration of pairwise neighbor topologies and miRNA family and cluster attributes for miRNA–disease association prediction. Brief. Bioinform. *23*, bbab428.
34. Ji, C., Gao, Z., Ma, X., Wu, Q., Ni, J., and Zheng, C. (2021). AEMDA: inferring miRNA–disease associations based on deep autoencoder. Bioinform *37*, 66–72.
35. Hu, X., Yin, Z., Zeng, Z., and Peng, Y. (2023). Prediction of miRNA–Disease Associations by Cascade Forest Model Based on Stacked Autoencoder. Molecules *28*, 5013.
36. Zhou, F., Yin, M.M., Jiao, C.N., Zhao, J.X., Zheng, C.H., and Liu, J.X. (2023). Predicting miRNA-disease associations through deep autoencoder with multiple kernel learning. IEEE Trans. Neural Netw. Learn. Syst. *34*, 5570–5579.
37. Liu, W., Lin, H., Huang, L., Peng, L., Tang, T., Zhao, Q., and Yang, L. (2022). Identification of miRNA–disease associations via deep forest ensemble learning based on autoencoder. Brief. Bioinform. *23*, bbac104.
38. Xie, X., Wang, Y., Sheng, N., Zhang, S., Cao, Y., and Fu, Y. (2022). Predicting miRNA-disease associations based on multi-view information fusion. Front. Genet. *13*, 979815.
39. Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J. Intern. Med. *4*, 627–635. https://pubmed.ncbi.nlm.nih.gov/24009950.
40. Saito, T., and Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One *10*, e0118432.
41. Yang, Z., Wu, L., Wang, A., Tang, W., Zhao, Y., Zhao, H., and Teschendorff, A.E. (2017). dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers. Nucleic Acids Res. *45*, D812–D818.
42. Xie, B., Ding, Q., Han, H., and Wu, D. (2013). miRCancer: a microRNA–cancer association database constructed by text mining on literature. Bioinform *29*, 638–644.
43. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. Adv. Neural Inf. Process. Syst. *30*, 5998–6008.
44. Huang, Z., Shi, J., Gao, Y., Cui, C., Zhang, S., Li, J., Zhou, Y., and Cui, Q. (2019). HMDD v3. 0: a database for experimentally supported human microRNA–disease associations. Nucleic Acids Res. *47*, D1013–D1017.
45. Kim, S., Yeganova, L., and Wilbur, W.J. (2016). Meshable: searching PubMed abstracts by utilizing MeSH and MeSH-derived topical terms. Bioinform *32*, 3044–3046.
46. Xuan, P., Chen, B., Zhang, T., and Yang, Y. (2021). Prediction of Drug–Target Interactions Based on Network Representation Learning and Ensemble Learning. IEEE/ACM Trans. Comput. Biol. Bioinform. *18*, 2671–2681.
47. Nair, V., and Hinton, G.E. (2010). Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th international conference on machine learning (ICML-10), pp. 807–814.
48. Zou, Y., Zhao, L., Qin, S., Pan, M., and Li, Z. (2020). Ship Target Detection and Identification Based on SSD_MobilenetV2. In 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC), pp. 1676–1680.
49. Bahdanau, D., Cho, K., and Bengio, Y. (2015). Neural Machine Translation by Jointly Learning to Align and Translate. In Proceedings of the 3rd International Conference on Learning Representations, 2015.
50. Kingma, D.P., and Ba, J. (2014). Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations, 2015.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
The version of miRNA-disease data	Database: HMDD v3.2	https://www.cuilab.cn/hmdd
Disease terms	The National Library of Medicine	https://www.nlm.nih.gov
Other		
Materials	This paper	N/A
Data and code	This paper	N/A

RESOURCE AVAILABILITY

Lead contact

Further requests for information should be directed and will be handled by the lead contact, Tiangang Zhang, email: zhang@hlju.edu.cn.

Materials availability

All materials reported in this paper will be shared by the lead contact upon request.

Data and code availability

- Data reported in this paper will be shared by the lead contact upon request.
- This paper does not report original code.
- Any additional information for reanalyzing this work is available from the lead contact upon request.

METHOD DETAILS

Given multiple types of miRNAs and diseases related through similarity data, our proposed CMMDA model for predicting miRNA-disease associations. Two miRNA-disease heterogenous networks are constructed according to one miRNA similarity, two disease similarities, and a miRNA-disease association. We encode feature vectors with the context relationships of the miRNA and disease nodes by transformer[43] for each heterogeneous network. Complementary information from each heterogeneous network is integrated through a proposed co-attention fusion mechanism to obtain complementary feature representations. The multi-modal attribute representations of the miRNA-disease node pairs are learned through a depthwise separable convolution-based module. These feature and attribute representations of miRNA-disease node pair are incorporated through a fully-connected layer, further evaluating association probability of node pairs.

Dataset

We collected known miRNA-disease associations from the HMDD database,[44] which includes 12,446 experimentally confirmed miRNA-disease associations from 853 miRNAs and 591 diseases. The disease terms are available at the U.S. National Library of Medicine.[45] Semantic similarities between two diseases are calculated based on representative directed acyclic graphs (DAGs)[14] composed of disease terms, and the functional similarities are calculated based on the associated miRNAs.

Multi-source data representations for miRNAs and diseases

Representations of miRNA-disease associations, miRNA similarities, and disease similarities

We define a matrix $A^{md} \in R^{N_m \times N_d}$ to denote the association between N_m miRNAs and N_d diseases. Two miRNAs with similar functions are more likely to be associated with similar diseases. So, if miRNA m_i is associated with disease d_j, then $A_{ij}^{md} = 1$. Otherwise, $A_{ij}^{md} = 0$. Separately, we derive the disease sets $\Phi_1 = \{d_1, d_5\}$ and $\Phi_2 = \{d_3, d_5, d_8\}$ related to miRNAs m_i and m_j, such that the similarity between Φ_1 and Φ_2 is considered the similarity between m_i and m_j. Therefore, the similarity among miRNAs is calculated utilizing the method proposed by Wang et al.[14] that forms a similarity matrix S^{mir}. A disease can be represented by a DAG structure that connects all semantic terms associated with the disease. If diseases d_i and d_j contain more common semantic terms, then they tend to be more similar. The method of Wang et al.[14] is applied again to measure semantic similarity of two diseases to obtain a matrix S^{dis_sem}. Similar diseases are more likely to interact with similar miRNAs. We specify the feature vector of disease d_i and d_j related to a miRNA to be h_i and h_j, respectively. The functional similarity of d_i and d_j is calculated as the cosine similarity[46] between h_i and h_j, and the functional similarity between all diseases forms the matrix S^{dis_fun}. A miRNA similarity matrix and two disease similarity matrices are expressed as,

$$S = \begin{cases} S^{mir} \in R^{N_m \times N_m} \\ S^{dis\_sem} \in R^{N_d \times N_d} \\ S^{dis\_fun} \in R^{N_d \times N_d} \end{cases} \quad \text{(Equation 1)}$$

where the values of each element in S^{mir}, S^{dis_sem}, and S^{dis_fun} range from 0 to 1, and a higher value indicates greater similarity between the corresponding nodes.

Multiple miRNA-disease heterogeneous networks

According to the similarity calculated for the two diseases, two bilayer miRNA-disease heterogeneous networks, defined as $G^{sem} = (V, E_{sem})$ and $G^{fun} = (V, E_{fun})$, are constructed. The node set $V = \{V^{mir} \cup V^{dis}\}$ is composed of the miRNA nodes V^{mir} and disease nodes V^{dis}. An edge $e_{ij} \in E$ exists if there is a connection between nodes $v_i \in V$ and $v_j \in V$. The attribute matrix for all miRNA and disease nodes is represented as,

$$X^{\Psi} = \begin{bmatrix} S^{mir} & A^{md} \\ A^{md^T} & S^{dis\_\Psi} \end{bmatrix}, X^{\Psi} \in R^{(N_m+N_d)\times(N_m+N_d)} \quad \text{(Equation 2)}$$

where A^{md^T} is the transpose matrix of A^{md} and ($\Psi = sem, fun$) represents that the disease similarity is computed according to semantic and functional perspectives of the disease.

Capturing context relationships and learning node complementary feature representations

Encoding node features with context relationships based on transformer

The miRNA and disease node attributes are important auxiliary information for identifying miRNA-disease association. Given the attribute matrices X^{sem} and X^{fun} of the miRNA and disease nodes, the i-th rows of each matrix record similarity between the i-th miRNA, m_i, and all other miRNAs, as well as the associations between m_i and all other diseases. The similarity of the $N_m + j$ disease, d_j, and all other diseases and associations with all miRNAs are represented in the $N_m + j$ row of X^{sem} and X^{fun}. For any heterogeneous network, the node attributes of a miRNA and its neighbors exist as context relationships.

We utilize the attribute information of the target and neighbor nodes to screen disease-related miRNAs. For X^{sem} and X^{fun}, we build two independent transformer encoders (TE^{sem} and TE^{fun}), each consisting of N_{layer} encoding layers. The goal is to learn the node feature with the context relationship between the miRNA and disease nodes. Because the encoding of X^{sem} and X^{fun} are similar, we present X^{sem} only in the following as an example to introduce the complete process.

Let the i-th miRNA node $m_i \in V$ be the target node and $v_j \in V$ be the neighbor miRNA and disease nodes. The context relationship between a target node and a neighbor node is aggregated based on the transformer. The feature vectors with d dimensions of m_i and v_j at the l-th coding layer are denoted as $p^l_{m_i}$ and $p^l_{v_j}$, respectively. Then, X^{sem} is passed into the first layer of the TE^{sem} as its input. The weight of the context relationship m_i to v_j is

$$\alpha^{h,l}_{(m_i,d_j)} = \frac{W^{h,l}_Q p^l_{m_i} \cdot W^{h,l}_K p^{l\ T}_{v_j}}{\sqrt{d}} \quad \text{(Equation 3)}$$

where h represents the h-th head attention and $W^{h,l}_Q$ and $W^{h,l}_K$ are the weight matrices of the head. The normalized attention weight is represented as,

$$\beta^{h,l}_{(m_i,v_j)} = \frac{\exp\left(\alpha^{h,l}_{(m_i,v_j)}\right)}{\sum_{v'_j \in U_{v_j}} \exp\left(\alpha^{h,l}_{(m_i,v'_j)}\right)} \quad \text{(Equation 4)}$$

where U_{v_j} is the set of neighboring nodes to m_i, and exp is the exponential function. The feature information of the neighbor nodes to m_i is aggregated to form $p^l_{m_i}$,

$$p^l_{m_i} = \left[\beta^{1,l}_{(m_i,v_j)} \cdot W^{1,l}_V p^l_{v_j} \parallel \beta^{2,l}_{(m_i,v_j)} \cdot W^{2,l}_V p^l_{v_j} \parallel \ldots \parallel \beta^{N_{head},l}_{(m_i,v_j)} \cdot W^{N_{head},l}_V p^l_{v_j}\right] \quad \text{(Equation 5)}$$

where N_{head} is the number of attention heads, W_V is the weight matrix, and $\parallel$ denotes the vector concatenation operation. The output of the last layer of TE^{sem} is P^{sem}, and $p^{sem}_{m_i}$ is regarded as a feature vector for m_i that integrates the neighbor node context relationship. Similarly, the TE^{fun} are passed to the N_{layer} encoding layers to obtain a feature matrix for all target miRNA and disease nodes. The obtained feature vector of m_i based on the disease functional similarity is represented as $p^{fun}_{m_i}$.

Co-attention fusion for complementing cross-network information

The constructed heterogeneous networks G^{sem} and G^{fun} defined above reflect the complex relationships among the miRNA and disease nodes and their complementarity. Therefore, the node feature vector $p^{sem}_{m_i}$ derived from G^{sem} and $p^{fun}_{m_i}$ from G^{fun} to enable the appropriate

interactions with the related information. We propose a co-attention fusion mechanism to obtain this complementary information. The importance score of $p_{m_i}^{fun}$ to $p_{m_i}^{sem}$ is defined as,

$$\varepsilon_{m_i}^{sem,fun} = Hp_{m_i}^{sem} \cdot Fp_{m_i}^{fun^T} \quad \text{(Equation 6)}$$

where H and F are weight matrices shared by the G^{sem} and G^{fun}, $p_{m_i}^{fun^T}$ is the transpose matrix of $p_{m_i}^{fun}$, and $\delta_{m_i}^{sem,fun}$ is the importance weight after normalization,

$$\delta_{m_i}^{sem,fun} = \frac{\exp\left(\varepsilon_{m_i}^{sem,fun}\right)}{\sum_{\varphi \in \{sem,fun\}} \exp\left(Hp_{m_i}^{sem} \cdot Fp_{m_i}^{\varphi^T}\right)} \quad \text{(Equation 7)}$$

Similarly, we calculate the normalized importance weights $\delta_{m_i}^{sem,sem}$, $\delta_{m_i}^{fun,sem}$, and $\delta_{m_i}^{fun,fun}$ of $p_{m_i}^{sem}$ to $p_{m_i}^{sem}$, $p_{m_i}^{sem}$ to $p_{m_i}^{fun}$, and $p_{m_i}^{fun}$ to $p_{m_i}^{fun}$, respectively. The $p_{m_i}^{sem}$ is input for a multi-layer perceptron (MLP), and the output feature vector after the interaction between $p_{m_i}^{sem}$ and $p_{m_i}^{fun}$ is

$$\tilde{p}_{m_i}^{sem} = MLP\left(\delta_{sem,sem}p_{m_i}^{sem} + \delta_{sem,fun}p_{m_i}^{fun}\right) \quad \text{(Equation 8)}$$

where the MLP contains two fully connected layers and a ReLU activation function.[47] We also derive the feature vector $\tilde{p}_{m_i}^{fun}$ with the interaction of $p_{m_i}^{fun}$ and $p_{m_i}^{sem}$, the latter of which stores the complementary information of node m_i in the G^{sem} network. So, we add $p_{m_i}^{sem}$ and $\tilde{p}_{m_i}^{sem}$ to form the complementary feature vector $\hat{p}_{m_i}^{sem}$ of m_i, such that

$$\hat{p}_{m_i}^{sem} = p_{m_i}^{sem} + \tilde{p}_{m_i}^{sem} \quad \text{(Equation 9)}$$

Following the same process outlined above, we obtain the complimentary feature vector $\hat{p}_{m_i}^{fun}$ of the node m_i corresponding to the G^{fun} network.

Integrating complementary features of multiple networks based on network-level attention

The complementary feature vectors $\hat{p}_{m_i}^{sem}$ and $\hat{p}_{m_i}^{fun}$ of m_i contribute differently toward predicting miRNA-disease associations, which is captured by constructing a score through a network-level attention mechanism,

$$s_{m_i}^{\Psi} = h^{net}\mathrm{ReLU}\left(W^{net}\hat{p}_{m_i}^{\Psi} + b^{net}\right) \quad \text{(Equation 10)}$$

where $\Psi \in \{sem,fun\}$, h^{net} and W^{net} are the weight vector and matrix, respectively, and b^{net} is a bias vector. The normalization of this attention weight is

$$\mu_{m_i}^{\Psi} = \frac{\exp\left(s_{m_i}^{\Psi}\right)}{\exp\left(s_{m_i}^{sem}\right)+\exp\left(s_{m_i}^{fun}\right)} \quad \text{(Equation 11)}$$

with a range in values of $[0, 1]$. Then, an attention-enhanced complementary feature representation for m_i is defined as,

$$z_{m_i} = \sum_{\Psi \in \{sem,fun\}} \left(\mu_{m_i}^{\Psi}\hat{p}_{m_i}^{\Psi} + \hat{p}_{m_i}^{\Psi}\right) \quad \text{(Equation 12)}$$

Similarly, we find the attention-enhanced complementary feature representation, z_{d_j}, for disease d_j.

Multi-modal attribute embedding of miRNA-disease node pairs

We leverage the hypothesis that a pair $m_i - d_j$ with an unknown association are more likely to be associated if they are related or similar to more common miRNAs and diseases. Considering the two types of drug similarities, an embedding method was proposed to obtain two modal attributes O^{sem} and O^{fun} of miRNA and disease node pairs. We use a miRNA m_i and disease d_j pair as an example to describe the process of building a representative embedding in the following. Given S^{mir}, S^{dis_sem}, and A^{md}, the i-th rows of A^{md} and S^{mir} are $(A^{md})_{i,*}$ and $(S^{mir})_{i,*}$, respectively, which are then spliced to form o_1^{sem},

$$o_1^{sem} = \left[(S^{mir})_{i,*} \,\|\, \left(A^{md}\right)_{i,*}\right], o_1^{sem} \in R^{(N_m+N_d)} \quad \text{(Equation 13)}$$

where $\|$ represents the concatenation operation, o_1^{sem} records the similarity between m_i and the other miRNA nodes as well as the associations between m_i and all the other disease nodes. The j-th rows $(A^{md})_{j,*}^T$ of A^{md^T} and $(S^{sem})_{j,*}$ of S^{dis_sem} are spliced to form o_2^{sem},

$$o_2^{sem} = \left[\left(A^{md}\right)_{j,*}^T \,\|\, \left(S^{dis\_sem}\right)_{j,*}\right], o_2^{sem} \in R^{(N_m+N_d)} \quad \text{(Equation 14)}$$

which is composed of the associations between d_j and all miRNA nodes and the similarities between d_j and all other disease nodes. Next, we stack o_1^{sem} and o_2^{sem} to acquire the first $m_i - d_j$ modal attribute embedding as,

$$O^{sem} = \begin{bmatrix} o_1^{sem} \\ o_2^{sem} \end{bmatrix}, O^{sem} \in R^{2 \times (N_m + N_d)} \quad \text{(Equation 15)}$$

Similarly, the second $m_i - d_j$ modal attribute embedding is obtained as,

$$O^{fun} = \begin{bmatrix} \left(S^{mir}\right)_{i,*} & \| & \left(A^{md}\right)_{i,*} \\ \left(A^{md}\right)_{j,*}^{T} & \| & \left(S^{dis\_fun}\right)_{j,*} \end{bmatrix}, O^{fun} \in R^{2 \times (N_m + N_d)} \quad \text{(Equation 16)}$$

These two modal attribute embeddings of $m_i - d_j$ are stacked to form the final multi-modal attribute embedding of

$$O^{final} = \left[O^{sem}; O^{fun}\right], O^{final} \in R^{2 \times 2 \times (N_m + N_d)} \quad \text{(Equation 17)}$$

where ; is an up-down stack operation.

Multi-modal attribute encoding by depthwise separable convolution

As defined above, the multi-modal attribute embeddings of the $m_i - d_j$ node pair is denoted as O^{final}, which are comprised of two channels of O_1^{final} and O_2^{final} that are extracted from G^{sem} and G^{fun}, respectively, so each has specific information. To better capture this distinction, we design a multi-modal attribute encoding module (MMAE) based on two depthwise separable convolution layers to learn O^{final}. This module consists of a depthwise convolution (DC) and pointwise convolution (PC), where DC applies a single filter on each input channel to capture its specific attribute information. The PC creates a linear combination of the DC layer outputs to capture the correlation between O_1^{final} and O_2^{final}.

To define the MMAE module, we first set $W_{dcon} \in R^{n_a \times 1 \times k_l \times k_w}$ as the DC filter, where n_a is the number of DC input channels, k_l and k_w are the length and width of the filter, respectively, and 1 is the depth of the filter. We apply a zero-padding operation to capture and learn the boundary information. A receptive field for the a-th single filter W_{dcon} slides along row m and column n of channel a is defined as,

$$Y_{a,m,n} = O_a^{final}(m: m + k_w, n: n + k_l) \quad \text{(Equation 18)}$$

where $m \in [1, 4 - k_w + 1], n \in [2 + N_m + N_d - k_l + 1], a \in [1, n_a]$. We obtain the element value $D_a(m, n)$ of the a-th feature map by applying the a-th single filter to $Y_{a,m,n}$ as,

$$D_a(m, n) = \sigma(W_{dcon}(a, :, :) * Y_{a,m,n}) \quad \text{(Equation 19)}$$

where $*$ indicates a convolution operation, σ and W_{dcon} are the ReLU6 activation function[48] and parameter matrix, respectively. The elemental values obtained from the n_a channels separately passing through the DC are stacked up and down to form $D = [D_1; D_2; \ldots; D_{n_a}]$. To integrate the information in the n_a channels, we send D into the PC layer, also called the 1×1 convolution layer. The c-th 1×1 filter is defined as $W_{pcon,c} \in R^{n_c \times n_a \times 1 \times 1}$, where n_c is the number of output channels of PC and 1 is the length and width of the filter. The element value U_c is obtained by performing the c-th 1×1 convolution operation on D, such that

$$U_c = W_{pcon,c} * D \quad \text{(Equation 20)}$$

where $c \in [1, n_c]$ and W_{pcon} is a trainable weight. U_c via max pooling layer forms the more representative attribute U_c'. Then, we derive the multi-modal attribute representation U_{moda} of the $m_i - d_j$ node pairs through two depthwise separable convolution layers and max pooling layers.

Final integration and optimization

The complementary feature representation of miRNA, m_i, and disease, d_j, are concatenated to achieve final complementary feature representation z_{ij} of $m_i - d_j$ node pairs as,

$$z_{ij} = \left[z_{v_i}^{mir} \| z_{v_j}^{dis}\right] \quad \text{(Equation 21)}$$

where $\|$ denotes a connection operation. The z_{ij} result is sent to a full-connection and softmax layers[49] to acquire association probability as,

$$\varsigma_{comp} = softmax\left(W_{comp} z_{ij} + b_{comp}\right) \quad \text{(Equation 22)}$$

where W_{comp} is a weight matrix and b_{comp} is a bias vector. The $\varsigma_{comp} = [\varsigma_{comp}^1, \varsigma_{comp}^2]$ represents an association probability distribution of two classes where ς_{comp}^1 and ς_{comp}^2 indicate the probability there is an association or not association between m_i and d_j, respectively. The cross-entropy loss function is calculated between the true label y_j and the predicted possibility score ς_{comp} as,

$$loss_{comp} = -\sum_{i=1}^{N_{train}}\sum_{j=1}^{C} y_j \log\left(\varsigma_{comp}^{j}\right) \quad \text{(Equation 23)}$$

where N_{train} is the number of training sample sets. The true label y_j indicates the actual association of a miRNA-disease node pairs, such that $y_j = 1$ when there is a real association between m_i and d_j, and $y_j = 0$ when no known relationship exists.

The multi-modal attribute representation U_{moda} is flattened as the vector u when input to the fully connected and softmax layers. The predicted association probability and cross-entropy loss are as follows,

$$\varsigma_{moda} = softmax(W_{moda}u + b_{moda}) \quad \text{(Equation 24)}$$

$$loss_{moda} = -\sum_{i=1}^{N_{train}}\sum_{j=1}^{C} y_j \log\left(\varsigma_{moda}^{j}\right) \quad \text{(Equation 25)}$$

where W_{moda} and b_{moda} denote trainable weight and bias vector, respectively. Then, $\varsigma_{moda} = [\varsigma_{moda}^{1}, \varsigma_{moda}^{2}]$ where ς_{moda}^{1} and ς_{moda}^{2} contain the related or unrelated probabilities of the nodes in the relationship $m_i - d_j$, respectively.

We optimize $loss_{comp}$ and $loss_{moda}$ with the Adam algorithm[50] and perform a weighted summation operation with ς_{comp} and ς_{moda} to obtain the final prediction score as,

$$\partial = \gamma \times \varsigma_{comp} + (1 - \gamma) \times \varsigma_{moda} \quad \text{(Equation 26)}$$

where $\gamma \in [0, 1]$ is a weighted coefficient that balances the contributions of the complementary feature and multi-modal attribute representations.

QUANTIFICATION AND STATISTICAL ANALYSIS

For each of prediction methods, we evaluated the average five-fold AUC (AUPR) result for the 591 diseases. When comparing two methods, we calculated the paired Wilcoxon test using 591 average five-fold AUC pairs (AUPR pairs). The statistical results in Table S4 indicate that our model consistently outperforms all other methods (p-value < 0.05).