



Article

Computational Method Using Attribute-Aware Message Passing and Graph Convolutional Network for Potential miRNA–Disease Association Prediction

Peng Qin ¹ and Jiyong An ^{2,*} 

¹ School of Computer Science, Liupanshui Normal University, Liupanshui 553004, China

² School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China

* Correspondence: ajy@cumt.edu.cn

Abstract

MicroRNA (miRNA) dysregulation is a crucial pathogenic factor that extensively participates in the occurrence and progression of various human diseases, especially cancers. Identifying unknown miRNA–disease connections is essential for understanding disease pathogenesis and improving clinical treatment strategies. Traditional biological experiments are often expensive and technically restricted, so computational prediction has become a widely used auxiliary research tool. In this study, we develop a novel predictive model called Attribute-Aware Message Passing Graph Convolutional Network (AAMPGCN) to identify potential miRNA–disease associations. The advantage of AAMPGCN lies in integrating miRNA and disease attribute information into the message-passing process: it partitions the miRNA–disease heterogeneous graph that incorporates miRNA functional similarity, disease semantic similarity, and Gaussian interaction kernel similarity into attribute-homogeneous subgraphs, while restricting high-order message propagation within each subgraph. This mechanism effectively filters cross-attribute noise, preserves the discriminability of miRNA and disease embeddings during deep convolution, and is thus well-adapted to miRNA–disease heterogeneous networks. The AAMPGCN prioritizes miRNA and disease attributes, aggregating messages specifically among nodes with similar attribute characteristics that are relevant to miRNA–disease interactions. Experimental results show that the AAMPGCN model achieves AUC and AUPR values of 94.06 and 93.52 on the HMDD2.0 dataset, which outperforms existing methods. The proposed AAMPGCN provides a new and effective method for miRNA–disease association prediction, and also offers theoretical support for the research on disease molecular mechanisms and the screening of clinical therapeutic targets.

Keywords: MDAs; attribute-aware message passing; GCN; multi-subgraph embedding fusion



Academic Editor: Richard H. W. Funk

Received: 16 May 2026

Revised: 1 July 2026

Accepted: 3 July 2026

Published: 7 July 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introductions

MicroRNAs (miRNAs) are a class of endogenous non-coding RNA molecules, typically 18–25 nucleotides in length, that play pivotal regulatory roles in a broad spectrum of cellular and molecular processes [1–4]. These processes include, but are not limited to, cell cycle progression, cellular differentiation, programmed cell death (apoptosis), and intercellular signal transduction [5–7]. Over the past decade, accumulating experimental evidence has demonstrated that perturbations in miRNA expression profiles or functional dysregulation

are closely correlated with the initiation, progression, and clinical prognosis of various human pathological conditions, including malignant tumors, cardiovascular disorders, and neurodegenerative diseases. By modulating the post-transcriptional expression of target genes, miRNAs are actively involved in the regulation of key biological pathways; their aberrant associations with diseases often disrupt the homeostasis of gene regulatory networks, ultimately contributing to the development and progression of pathological phenotypes [8–11].

Identifying the intricate associations between miRNAs and diseases is fundamental to deciphering the molecular mechanisms underlying disease pathogenesis, developing sensitive early diagnostic biomarkers, and identifying potential therapeutic targets for precision medicine. Nevertheless, conventional experimental techniques employed to validate miRNA–disease associations—such as luciferase reporter assays, PCR and RNA sequencing are inherently associated with high costs, prolonged experimental cycles, and intensive labor requirements [12–15]. These inherent limitations pose a significant challenge in keeping pace with the exponential growth of miRNA and disease-related omics data, leading to a substantial number of unconfirmed potential miRNA–disease associations that remain unexplored and require further investigation [16,17]. To address this bottleneck, computational prediction methods have been extensively developed and applied as an effective complement to traditional experimental verification, greatly accelerating the discovery of potential miRNA–disease associations.

At present, the existing computational prediction methods for miRNA–disease associations can be roughly divided into several categories, including similarity-based methods, matrix completion-based methods, machine learning-based methods, and graph neural network-based methods, each with its own unique advantages and inherent deficiencies [18–20]. Similarity-based methods are simple and interpretable, enabling rapid preliminary screening without complex training, but they often rely on single-type similarity information, suffer from inaccurate predictions when similarity data is sparse, and fail to capture complex non-linear relationships. For instance, Cui et al. [21] proposed a method to construct pairwise functional similarities and functional networks of human miRNAs via disease association structures. Comparative results demonstrate that the obtained miRNA functional similarities are consistent with known miRNA functional links. Chen et al. [12] proposed three miRNA–disease association prediction approaches: MBSI (microRNA-based similarity inference), PBSI (phenotype-based similarity inference), and NetCBI (network-consistency-based inference). All three leverage global network similarity metrics to identify unreported miRNA–disease pairs. Chen et al. [22] developed a WBSMDA (Within and Between Score for MiRNA–Disease Association prediction) to identify potential miRNA–complex disease associations. This model works for diseases lacking known relevant miRNAs, and leave-one-out cross-validation yields an AUC of 0.8031, verifying its stable predictive capacity. Le et al. [23] adopt a customized random walk with restart algorithm to infer disease–miRNA associations, utilizing complementary information from validated/predicted miRNA–target interactions and disease phenotypic similarities. Matrix completion-based methods excel at handling sparse association matrices and mining hidden patterns, yet they have high computational complexity, struggle with cold-start problems, and ignore node attribute differences. For instance, Chen et al. [24] propose IMCMDA (Inductive Matrix Completion for MiRNA–Disease Association prediction). We compute integrated miRNA and disease similarities from miRNA functional similarity, disease semantic similarity, and Gaussian interaction profile kernel similarity. The model recovers unobserved miRNA–disease associations using known links and the above integrated similarities. Chen et al. [25] propose LRMCMDA, a low-rank matrix completion method for miRNA–disease association prediction. It first builds a bipartite miRNA–disease

graph from known associations and constructs an R-projected miRNA graph where edges connect miRNAs sharing common disease neighbors. A corresponding D-projected disease graph is defined analogously. Li et al. [26] construct AMCMDA (Accurate Matrix Completion Model) for latent miRNA–disease association prediction, which improves prediction accuracy via truncated nuclear norm minimization. AMCMDA first builds a heterogeneous network integrating miRNA similarities, disease similarities and known associations. Next, an optimization framework is devised to approximate the truncated nuclear norm and recover missing entries of the association matrix. Finally, the alternating direction method of multipliers (ADMMs) is adopted to solve the optimization and output prediction scores. Qin et al. [27] propose EMCMDA (Efficient Matrix Completion for MiRNA–Disease Association prediction) to infer miRNA–disease associations. First, we compute multi-source similarities for miRNAs and diseases and fuse them into unified similarity metrics. Second, we construct a heterogeneous network with these biological similarities and build the target association matrix. Finally, we formulate association prediction as a low-rank matrix completion problem solved by minimizing the truncated Schatten p-norm. Traditional machine learning methods exhibit good adaptability to different data types and possess certain generalization ability. Traditional machine learning-based methods adapt well to different data types and have certain generalization ability. For instance, Wei L. et al. [28] proposed a novel DFELMDA model that first constructs multi-type feature representations for miRNAs and diseases, and adopts two deep autoencoders to extract low-dimensional latent features. Deep random forest is then used to generate and fuse prediction scores for final inference. Chen X. et al. [29] propose EGBMMDA, an extreme gradient boosting machine model for miRNA–disease association prediction and the first decision tree learning-based model for such a prediction task. Qiguo D. et al. [30] present ERMDA, a resampling-based ensemble framework for miRNA–disease association prediction. It constructs balanced training subsets to alleviate sample imbalance and extracts miRNA and disease features using multi-source similarity and verified association information. Kai Z. et al. [31] present a machine learning model called MLMMDA that uses a random forest classifier for miRNA–disease association prediction. Graph neural network-based methods effectively capture complex network association patterns and integrate topological and attribute information. For instance, Wang S. et al. [32] propose a heterogeneous graph attention network based on meta-subgraphs called MSHGANMDA that first define five meta-subgraph types from known associations, then use meta-subgraph attention and semantic attention to extract features of miRNA–disease pairs within and between these meta-subgraphs, and, finally, apply a fully connected layer to predict association scores and cross-entropy loss for end-to-end training. Han H. et al. [33] adopts graph convolutional networks to learn feature embeddings of miRNAs and diseases, and introduces an attention mechanism to fuse embeddings output from multiple convolutional layers. Finally, based on the fused embeddings, scores are predicted for unknown miRNA–disease associations. Gao S. et al. [34] proposed a GCN-based model DEJKMDR for miRNA–disease association prediction, integrating biomolecular information and Gaussian interaction profile kernel similarity to avoid overfitting, with adaptive weight allocation for feature fusion and accurate association scoring. Liang X. et al. [35] proposed a novel method for miRNA–disease association prediction, which integrates a graph convolutional network and a hypergraph convolutional network. The graph convolutional network extracts feature from miRNA and disease similarity data, while the hypergraph convolutional network captures complex high-order interactions of known miRNA–disease associations based on the learned representations. In summary, although existing computational methods for miRNA–disease association prediction have their respective advantages and have achieved remarkable progress in assisting biological research, there is still room for improvement in feature

utilization, computational efficiency, and sparse data processing. Therefore, it is urgent to develop a novel computational model that can fully integrate multi-source information and improve prediction accuracy and robustness.

In this study, we develop a novel predictive model called Attribute-Aware Message Passing Graph Convolutional Network (AAMPGCN) to identify potential miRNA–disease associations. The advantage of AAMPGCN lies in integrating miRNA and disease attribute information into the message-passing process: it partitions the miRNA–disease heterogeneous graph that incorporates miRNA functional similarity, disease semantic similarity, and Gaussian interaction kernel similarity into attribute-homogeneous subgraphs, while restricting high-order message propagation within each subgraph. This mechanism effectively filters cross-attribute noise, preserves the discriminability of miRNA and disease embeddings during deep convolution, and is thus well-adapted to miRNA–disease heterogeneous networks. The AAMPGCN prioritizes miRNA and disease attributes, aggregating messages specifically among nodes with similar attribute characteristics that are relevant to miRNA–disease interactions. Experimental results show that the AAMPGCN model achieves AUC and AUPR values of 94.06 and 93.52 on the HMDD2.0 dataset, which outperforms existing methods. The proposed AAMPGCN provides a new and effective method for miRNA–disease association prediction, and also offers theoretical support for the research on disease molecular mechanisms and the screening of clinical therapeutic targets.

2. Results

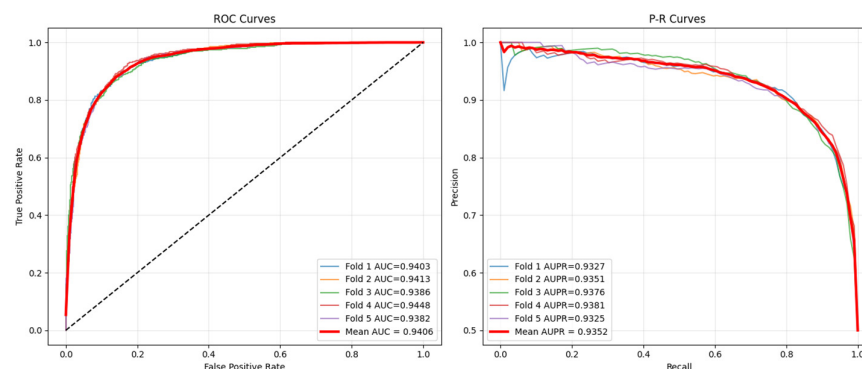
2.1. Performance of the Proposed Method

To fully assess the effectiveness of the proposed AAMPGCN model in predicting miRNA–disease associations (MDAs), we carried out strict experiments on the HMDD v2.0 dataset, which is widely used in MDA prediction research and contains 495 miRNAs, 383 diseases, and 5430 known miRNA–disease associations. To maintain class balance, an equal number of negative samples was randomly extracted from unlabeled association pairs. The 5-fold cross-validation (5-CV) method was employed to ensure the credibility of the experimental results and avoid overfitting.

The proposed AAMPGCN model was built based on PyTorch 2.0 and trained for over 1000 epochs. After parameter tuning, the learning rate was set to 0.0006 with a dropout rate of 0.4. Several classic classification metrics were adopted to assess model performance, and the corresponding quantitative results and ROC, as well as PR curves, are presented in Table 1 and Figure 1, respectively. Figure 1 shows the Receiver Operating Characteristic (ROC) curves and Precision–Recall (PR) curves plotted in this work, which quantitatively characterize the predictive performance of our proposed model. The ROC curve plots true positive rate (TPR) against false positive rate (FPR) under varied classification thresholds. The area enclosed by the ROC curve (AUC) reflects the model's general discrimination capacity: an AUC close to 1 implies nearly perfect classification, whereas an AUC of 0.5 equals random guesses. Considering the severe class imbalance inherent in miRNA–disease datasets, ROC evaluation alone may overestimate model performance. Therefore, we further adopted PR curves as a complementary assessment metric. The PR curve depicts Precision versus Recall, and the corresponding AUPR value provides a more reliable evaluation for imbalanced biological data. Figure 1 illustrates that our proposed approach obtains an average AUC value of 94.06 on the ROC curve and an average AUPR value of 93.52 on the PR curve, exceeding all baseline models, which further demonstrates the reliability and superiority of the model developed in this work.

Table 1. AAMPGCN evaluation on HMDD v2.0 via five-fold cross-validation.

Testing Set	Acc. (%)	Rec. (%)	Pre. (%)	F1 (%)
1	86.97	88.21	86.67	87.13
2	87.11	89.13	85.66	87.36
3	86.62	88.12	85.82	86.65
4	87.48	89.78	86.83	87.76
5	86.93	86.92	86.76	86.84
Mean	87.02 ± 0.31	88.43 ± 1.09	86.35 ± 0.56	87.15 ± 0.43

**Figure 1.** ROC and P-R plots for AAMPGCN trained on HMDD v2.0.

2.2. Comparison with Other Models

To further explore the prediction performance of AAMPGCN in identifying miRNA–disease associations (MDAs), we compared our method with several state-of-the-art computational models:

1. HHOMR [36]: This model exploits high-order statistical characteristics and graph structural information, and adopts an attention mechanism to effectively mine implicit correlation patterns between miRNAs and diseases.
2. MTLMDA [37]: It integrates miRNA–disease and gene–disease networks through cross-layer compression, employs graph convolutional networks to extract node features, and utilizes a linear decoder to complete association prediction.
3. VGAEMDA [38]: The model leverages the variational graph autoencoder framework to conduct unsupervised representation learning, thereby capturing underlying miRNA–disease association relationships.
4. GBDT-LR [39]: It first adopts matrix factorization to acquire low-dimensional embedding representations of the miRNA–disease adjacency matrix, and then excavates potential association dependencies under a label space learning paradigm.

Comparisons among the aforementioned models and our developed model are summarized in Table 2 and Figure 2.

Table 2. Comparative evaluation of model performance.

Model	AUC (%)	Acc. (%)	Rec. (%)	Pre. (%)	F1 (%)
HHOMR	93.28	86.25	—	85.54	85.68
MTLMDA	93.05	85.62	84.60	84.81	84.58
VGAEMDA	93.85	87.13	91.03	84.85	87.60
GBDT-LR	92.74	83.04	89.31	83.15	83.02
AAMPGCN	94.06	87.02	88.43	86.35	87.15

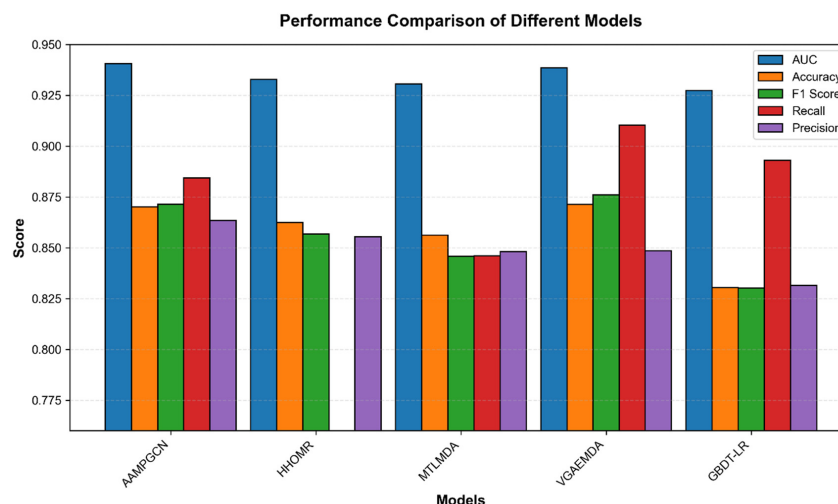


Figure 2. Bar Chart for Comparative Evaluation of Model Performance.

As illustrated in Figure 2 and Table 2, our model outperforms all competing methods, mainly for the following reasons: the AAMP GCN integrates three biologically meaningful similarity matrices into a unified graph learning framework. The AAMP GCN integrates miRNA and disease attribute information into the message-passing process, partitioning attribute-homogeneous subgraphs, and restricting high-order message propagation within subgraphs, which effectively filters cross-attribute noise, avoids over-smoothing, and maintains the discriminability of node embeddings. Moreover, the AAMP GCN prioritizes miRNA and disease attributes, aggregating messages specifically among nodes with similar attribute characteristics that are relevant to miRNA–disease interactions. This is the core reason why AMPGCN outperforms other state-of-the-art methods in MDA prediction.

2.3. Hyperparameter Sensitivity Analysis

To improve the prediction performance of the proposed AAMP GCN model and ensure its stability and generalization ability, hyperparameter optimization experiments were carried out, focusing on two key hyperparameters that significantly influence the model's feature learning and representation capabilities: the polynomial order K of graph convolutional layers and the hidden layer dimension. All experiments were implemented based on the PyTorch framework, with consistent training settings (1000 epochs, learning rate of 0.0006, dropout rate of 0.4, and Adam optimizer with weight decay of 1×10^{-4}) to isolate the impact of the two target hyperparameters on model performance. The model's performance was evaluated using the AUC and AUPR values as the primary metric to ensure the comprehensiveness and reliability of the evaluation results. The polynomial order K is a crucial hyperparameter that determines the scope of local neighborhood aggregation during the feature learning process. A smaller K value may lead to insufficient extraction of neighborhood information, resulting in underfitting and limited model capacity; on the contrary, an excessively large K value may introduce redundant noise from distant nodes, increase model variance, and further cause overfitting. To determine the optimal value of K , a grid search was performed within the range of 1 to 5. Experimental results demonstrated that the model achieved the best comprehensive performance when $K = 3$, with an average AUC of 0.9415. When $K = 1$ and $K = 2$, the model only aggregated information from directly adjacent nodes, leading to inadequate feature learning and a relatively low AUC and AUPR. For $K \geq 3$, the model aggregated information from more distant nodes, which resulted in increased noise interference and a slight decrease in AUC and AUPR for $K = 4$ and $K = 5$. This confirms that $K = 3$ can achieve an optimal balance between model capacity and stability.

The hidden layer feature dimension is another core hyperparameter that affects the model's ability to learn high-dimensional feature representations of miRNA and disease nodes. A too-small feature dimension may restrict the model's feature learning capacity, making it difficult to capture the complex latent relationships between miRNAs and diseases; a too-large feature dimension may increase the model's computational complexity, induce overfitting, and reduce the model's generalization ability. A grid search was conducted for the feature dimension within the range of 32, 64, 128, and 256, which are widely used in graph neural network models for miRNA–disease association prediction. Experimental results showed that the model achieved the highest AUC of 0.9421 when the feature dimension was set to 128. When the feature dimension was 32 and 64, the model's feature representation ability was insufficient, resulting in a low AUC. When the feature dimension was increased to 256, although the model's feature learning capacity was enhanced, it led to overfitting, and the AUC decreased to 0.8636. Therefore, 128 was determined as the optimal feature dimension for AAMPGCN, which can effectively balance the feature representation ability and computational efficiency.

In summary, the optimal hyperparameters of AAMPGCN determined through the above experiments are as follows: the polynomial order $K = 3$ of graph convolutional layers and the hidden layer dimension = 128. These optimal hyperparameters enable the model to effectively aggregate local neighborhood information, learn discriminative feature representations, and thus achieve excellent performance in miRNA–disease association prediction. The comparison of AUC and AUPR under different values of the two hyperparameters is presented in Figures 3 and 4.

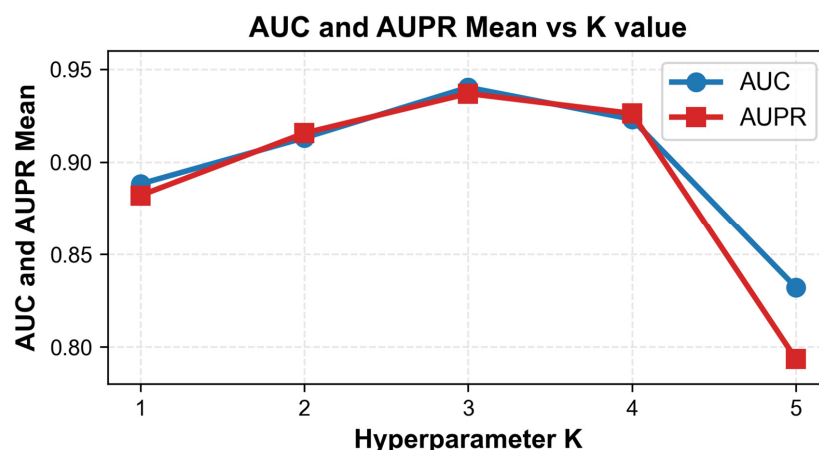


Figure 3. Result of employing different k values in AAMPGCN.

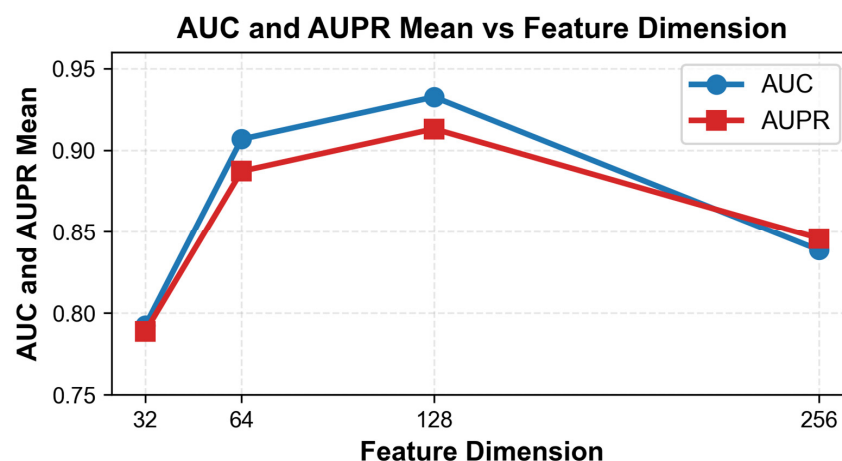


Figure 4. Result of employing different feature dimensions in AAMPGCN.

2.4. Case Study

The AAMPGCN method was used to test the accuracy of predictive associations via three cases (prostate, lung, and pancreatic cancer), with predicted candidate miRNAs verified against the DBDEMC [40] and Phenomir databases [41].

In the first case, prostate neoplasms, which are the most common male malignancy globally and cause over 100,000 European male deaths in 2018 [42], were selected to test AAMPGCN's applicability to novel diseases without known associated miRNAs. All miRNA–disease associations for prostate neoplasms in HMDD 2.0 were set to zero, followed by the application of AAMPGCN to identify associated miRNAs. Figure 5 lists AAMPGCN's top 48 candidate miRNAs for prostate neoplasms, all of which were verified by the DBDEMC and Phenomir databases. Notably, the first-ranked hsa-miR-96b regulates prostate cancer cell apoptosis by inhibiting FoxO1, thereby validating AAMPGCN's predictive ability for such novel diseases. For the second case, we conducted a study on lung neoplasms, which is a fatal global cancer with a low five-year survival rate [43] and a disease where miRNAs assist in diagnosis and treatment. We adopted the same data processing method as that used in the first case. Among AAMPGCN's top 46 predicted miRNAs for lung neoplasms listed in Figure 6, 46 were validated by one or both of the two databases. Ectopic expression of miR-494 in A549 lung cancer cells promoted tumor-initiating cells, enhanced cell motility, and upregulated the expression of stem cell-related genes, supporting AAMPGCN's value for the diagnosis and treatment of lung cancer. For the third case, pancreatic tumors were selected as the new disease, and the same experimental method as that used in the previous two cases was adopted. The top 49 predicted results are shown in Figure 7, all of which have been verified by one or both of the two databases. Existing studies have shown that increased serum miR-193b expression can serve as a potential new biomarker for pancreatic neuroendocrine tumors (PNEN).

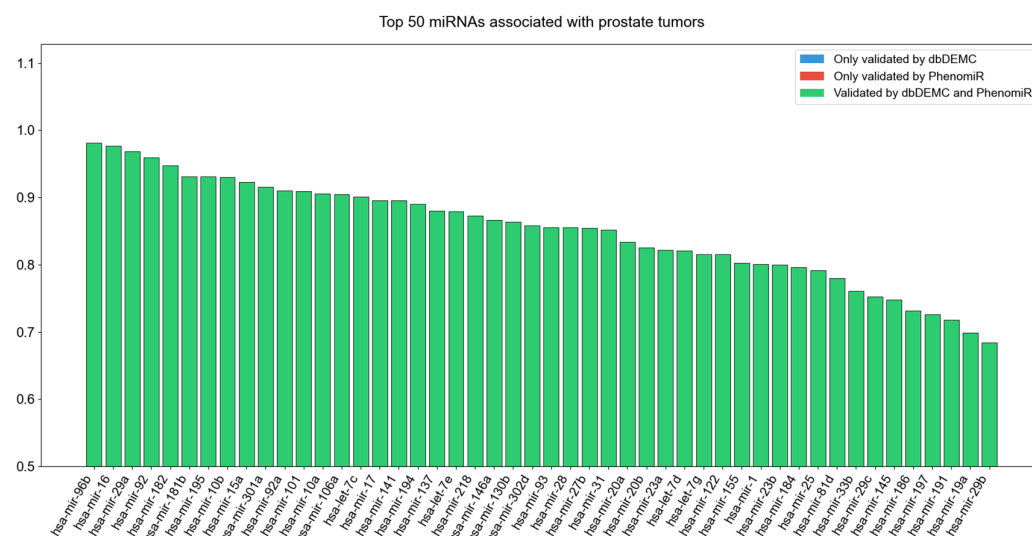


Figure 5. Prediction of top 50 miRNAs related to prostate tumors.

The experimental results of the above three cases further demonstrate the important role of the AAMPGCN model in the prediction of new diseases.

To further verify the effectiveness of AAMPGCN, we compared the counts of database-validated miRNA–disease associations among the top 50 prediction candidates generated by our model and other competing baselines. Specifically, we performed this statistical comparison on two representative diseases covered in this work, namely lung cancer and prostate cancer, by tallying the total number of biologically confirmed miRNA–disease

pairs ranked within the top 50 predictions for each competing method. All comparative outcomes are summarized in Table 3.

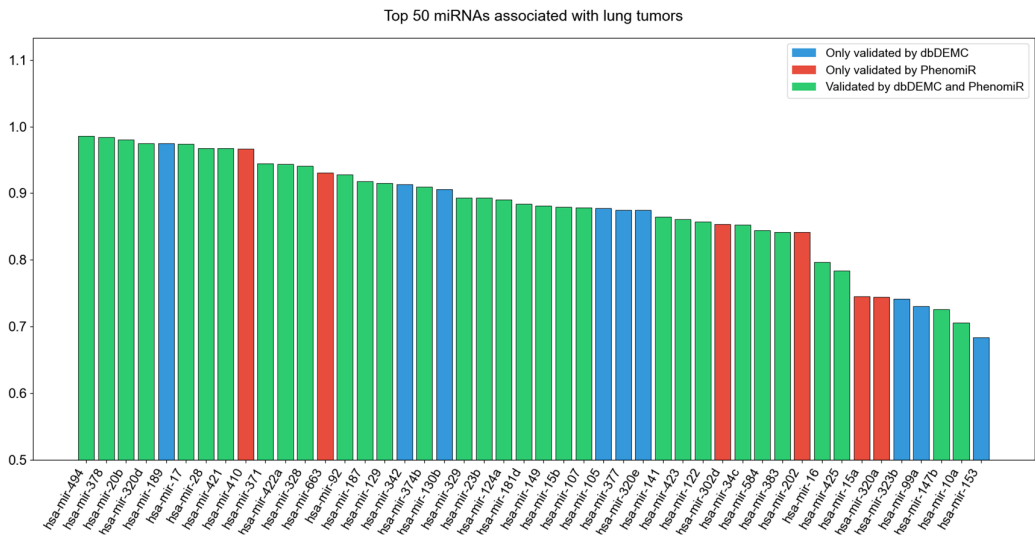


Figure 6. Prediction of top 50 miRNAs related to lung tumors.

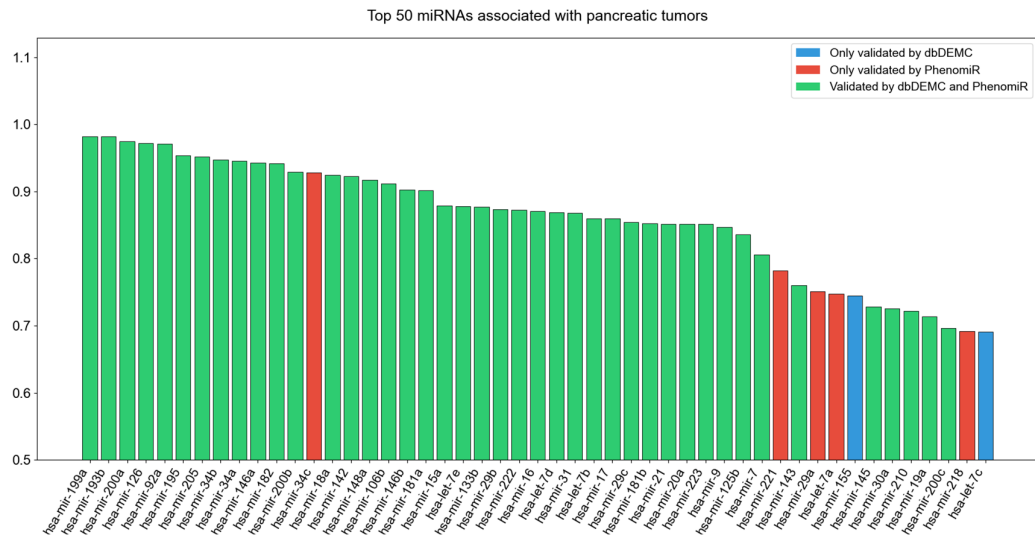


Figure 7. Prediction of top 50 miRNAs related to pancreatic tumors.

Table 3. The total number of database-validated miRNA–disease associations among the top 50 predicted entries of each prediction model.

Model	Lung Tumors	Prostate Tumors
HOMR	37	43
GCCDN [44]	-	47
GBDT-LR	-	44
AAMPGCN	46	49

As shown in Table 3, AAMPGCN identifies more validated miRNA–disease associations than all competing prediction models, further verifying its superior predictive performance. This advantage mainly stems from the unique architecture of AAMPGCN. The model incorporates attribute features of miRNAs and diseases into the message-passing scheme to build a high-quality heterogeneous graph fused with multiple similarity measurements. Meanwhile, it constrains high-order message propagation within attribute-homogeneous subgraphs. This tailored design effectively eliminates cross-attribute noise

and mitigates the ubiquitous over-smoothing problem of traditional graph convolutional networks. Consequently, high-discriminative node embeddings are learned, enabling the model to precisely capture inherent biological correlations between miRNAs and diseases.

3. Discussion

Our proposed Attribute-Aware Message Passing Graph Convolutional Network (AAMPGCN) demonstrates superior performance in MDA prediction, which is fully validated by systematic experiments and case studies. Experiments were conducted on the HMDD 2.0 dataset and used a 5-fold cross-validation strategy to ensure the reliability and generalizability of the results and to avoid overfitting. Key evaluation metrics included Area Under the Receiver Operating Characteristic Curve (AUC), Area Under the Precision–Recall Curve (AUPR), Accuracy (ACC), Precision, Recall, and F1-score, among which AUC and AUPR were the primary indicators to reflect the model’s predictive ability. AAMPGCN achieved excellent performance across all metrics: an AUC of 0.9406, an AUPR of 0.9352, an ACC of 0.8702, an F1-score of 0.8715, and a precision of 0.8635. This performance significantly outperformed mainstream baseline models, including HHOMR, MTLMDA, VGAEMDA, and GBDT-LR.

The excellent experimental performance of AAMPGCN can be attributed to its core design features, which are specifically tailored to the characteristics of miRNA–disease heterogeneous networks. Specifically, AAMPGCN integrates miRNA and disease attribute information into the message-passing process, constructs a heterogeneous graph incorporating miRNA functional similarity, disease semantic similarity, and Gaussian interaction kernel similarity, and restricts high-order message propagation to attribute-homogeneous sub-graphs. This design effectively filters out cross-attribute noise, avoids the over-smoothing problem that plagues traditional graph convolutional networks, and ensures the generation of high-quality node embeddings, thereby enabling the model to accurately capture the intrinsic interactions between miRNAs and diseases.

Case studies on prostate, lung, and pancreatic cancer—three common malignant tumors closely associated with miRNA dysregulation—further validated the practical value of AAMPGCN. Among the top 50 predicted MDAs for each disease, all were supported by authoritative databases or the published experimental literature.

In summary, AAMPGCN’s superior experimental performance confirms its reliability as a powerful computational tool for MDA research, providing valuable support for subsequent biomedical experiments and clinical research.

Despite AAMPGCN’s promising performance in miRNA–disease association (MDA) prediction, it has several limitations to address in future work:

- Dependence on similarity matrix quality: Its performance is closely linked to the quality and coverage of input similarity matrices, but incomplete annotations in existing biological databases (e.g., insufficient information for rare miRNAs/diseases) lead to inaccurate similarity calculations and compromised performance.
- Lack of regulatory pathway modeling: It focuses on pairwise miRNA–disease associations but fails to explicitly model intermediate regulatory mechanisms (e.g., miRNA–gene–disease networks), which limits model interpretability.
- Transudative learning limitation: As a transudative model, it can only predict for miRNAs/diseases in the training set and cannot generalize to unseen nodes without retraining, restricting practical application in emerging disease research.

To address these issues, future work will focus on three directions: (1) Integrate multi-omics data to construct more comprehensive and accurate miRNA/disease similarity metrics, reducing the impact of incomplete annotations. (2) Design an inductive graph learning module (via meta-learning or contrastive learning) to enable generalization to unseen

miRNAs/diseases without retraining. (3) Incorporate pathway-level attention mechanisms to model miRNA–gene–disease regulatory pathways, enhancing model interpretability and providing deeper mechanistic insights.

4. Materials and Methods

4.1. Datasets

The Human microRNA Disease Database (HMDD2.0) is a well-known benchmark database that collects experimentally verified miRNA–disease associations. The HMDD2.0 includes 495 miRNAs, 383 diseases, and 5430 confirmed associations [45]. To facilitate subsequent analysis, we constructed an adjacency matrix $A \in \mathbb{R}^{n_d \times n_m}$ to formalize the associations between miRNAs and diseases. Here, n_m and n_d denote the number of known miRNAs and known diseases, respectively. Specifically, if the disease d_i has been experimentally verified to be associated with the miRNA m_j , the element A_{ij} is set to 1; otherwise, it is 0.

4.2. miRNA Functional Similarity

Based on the hypothesis that miRNAs with analogous functions tend to be associated with phenotypically similar diseases [21]. In this study, the functional similarity information for miRNAs FS , obtained from the MISIM (miRNA similarity) resource, we used matrix FS (miRNA functional similarity) to characterize miRNA functional similarity, where the entry $FS(m_i, m_j)$ denotes the similarity score between miRNA m_i and miRNA m_j . The final FS value is calculated via cosine similarity, as shown in Equation (1):

$$FS(m_1, m_2) = COS_{mir}(m_1, m_2) = \frac{f_{m1} \cdot f_{m2}^T}{||f_{m1}|| \cdot ||f_{m2}||} \quad (1)$$

where f_{m1} and f_{m2} represent enhanced disease semantic feature vectors of miRNA m_1 and miRNA m_2 , respectively; $f_{m1} \cdot f_{m2}^T$ represents the dot product of two vectors and $||f||$ represents the L2 norm of the vector.

4.3. Disease Semantic Similarity

Disease semantic similarity is established relying on the Medical Subject Headings (MeSH) database [46–48]. The internal connections among various diseases can be intuitively abstracted through a Directed Acyclic Graph (DAG). Specifically, the topological structure of disease D is defined as $DAG(D) = (T(D), E(D))$. In this definition, $T(D)$ refers to the node collection covering the target disease D and all its ancestral nodes, while $E(D)$ stands for the edge set connecting these associated nodes. For an arbitrary disease node d contained in $DAG(D)$, the semantic contribution is as follows [49]:

$$D1_D(d) = \begin{cases} 1 & \text{if } d = D \\ \max\{\Delta \cdot D1_D(d') \mid d' \in \text{children of } d\}, & \text{if } d \neq D, \end{cases} \quad (2)$$

where Δ represents the semantic contribution decay factor, usually set to 0.5 [47]. The semantic similarity of disease D is defined as

$$DS1(D) = \sum_{d \in T(D)} D1_D(d) \quad (3)$$

The semantic contribution score described above is represented by $SS1$, where each element $SS1(i, j)$ is computed via the following formula:

$$SS1(i, j) = \frac{\sum_{t \in T(i) \cap T(j)} (D1_i(t) + D1_j(t))}{DS1(i) + DS1(j)} \quad (4)$$

From Equations (1) and (2), it can be observed that within the DAG structure of disease D , the semantic contribution of the root node D itself is defined as 1. As the topological distance from D to its ancestor nodes increases, the semantic contribution of these nodes to D decays progressively. It is also clear that nodes located at the same hierarchical level share identical semantic contribution scores. Considering that nodes unique to $DAG(D)$ tend to contribute more to the semantics of disease D than those shared across multiple disease DAGs, Xuan et al. [37] developed an alternative computational model, as described below.

$$DS2(D) = -\log\left(\frac{\text{the numbers of DAGs includ } d}{\text{the numbers of diseases}}\right) \quad (5)$$

Correspondingly, the alternative semantic contribution measure is formulated as follows:

$$DS2(D) = \sum_{d \in T(D)} D2_D(d) \quad (6)$$

$$SS2(i, j) = \frac{\sum_{t \in T(i) \cap T(j)} (D2_i(t) + D2_j(t))}{DS2(i) + DS2(j)} \quad (7)$$

In conclusion, the disease semantic similarity is defined by the formula presented below:

$$SS(i, j) = \frac{ss1(i, j) + ss2(i, j)}{2} \quad (8)$$

4.4. Gaussian Interaction Profile Kernel Similarity

As a widely adopted measurement for assessing entity similarity, the Gaussian interaction profile (GIP) kernel similarity is extensively applied in relevant research. The GIP kernel for miRNAs can be computed via Equations (2)–(8) [50–52]:

$$GM(i, j) = \exp\left(-\gamma_m \|A(m_i) - A(m_j)\|^2\right)$$

in which $A(m_i)$ refers to the association profile of miRNA i , while $A(m_j)$ corresponds to that of miRNA j . γ_m serves as a tunable parameter for kernel bandwidth optimization, with the specific calculation equation shown below:

$$\gamma_m = \frac{\gamma'_m}{\left(\frac{1}{nm} \sum_{i=1}^{nm} \|A(m_i)\|^2\right)} \quad (9)$$

in which γ'_m is assigned a fixed value of 1. Likewise, the Gaussian interaction profile kernel for diseases is formulated as

$$GD(i, j) = \exp\left(-\gamma_d \|A(d_i) - A(d_j)\|^2\right) \quad (10)$$

$$\gamma_d = \frac{\gamma'_d}{\left(\frac{1}{nd} \sum_{i=1}^{nd} \|A(d_i)\|^2\right)} \quad (11)$$

in which $A(d_i)$ and $A(d_j)$ denote the interaction profiles corresponding to disease i and disease j , with γ'_d assigned as 1.

4.5. Joint Similarity Kernel of miRNAs and Diseases

In this work, we construct a miRNA–disease joint similarity kernel to fuse multiple biological similarity information. Concretely, by combining the functional similarity kernel (FS) and the Gaussian interaction profile kernel (GM) of miRNAs, the comprehensive

miRNA similarity kernel SM can be established. The definition of this fused similarity kernel is given below:

$$SM = \begin{cases} FS(i, j), & FS(i, j) \neq 0 \\ GM(i, j), & \text{otherwise} \end{cases} \quad (12)$$

According to the above definition, the Gaussian interaction profile kernel GM is utilized to fill the missing values of the functional similarity matrix FS in the fused similarity kernel. In a similar manner, the disease similarity kernel SD is formulated as

$$SD = \begin{cases} SS(i, j), & SS(i, j) \neq 0 \\ GD(i, j), & \text{otherwise} \end{cases} \quad (13)$$

4.6. Attribute-Aware Message Passing GCN(AAMPGCN)

The Attribute-aware Message-Passing Graph Convolutional Network (AAMPGCN), proposed in 2025 [14,53], is an improved graph convolutional model designed to address the over-smoothing problem and inefficient attribute utilization in traditional Graph Convolutional Networks (GCNs). Different from traditional GCNs that only rely on graph structural information for global message passing, AAMPGCN emphasizes the guiding role of node attributes. It assumes that nodes with similar attribute characteristics tend to have similar semantic meanings and interaction patterns, and, thus, high-order message aggregation should be carried out within the scope of these similar nodes. The entire theoretical framework of AAMPGCN consists of three key components: attribute-aware subgraph partitioning, constrained message passing within subgraphs, and multi-subgraph embedding fusion. These components work together to achieve effective integration of node structural information and attribute information, thereby improving the performance of graph learning tasks. This mechanism not only filters out heterogeneous noise caused by cross-attribute propagation but also maintains the discriminability of node embeddings during deep convolution, making it particularly suitable for sparse and heterogeneous networks such as biological information networks. The flowchart of the AAMPGCN model for predicting MDAs is shown in Figure 8.

4.6.1. Attribute-Aware Subgraph Partitioning

To facilitate the derivation of formulas, the following definitions and symbols are specified for the graph structure and node characteristics involved in AAMPGCN:

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$ denote the input graph, where $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$ is the set of nodes (with N being the total number of nodes), $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges representing the connections between nodes, and $\mathbf{X} \in \mathbb{R}^{N \times D}$ is the node attribute matrix (with D being the dimension of node attributes).

Let $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_K\}$ represent the set of attribute-homogeneous subgraphs obtained by attribute clustering, where K is the number of subgraphs, $\mathcal{S}_k \subseteq \mathcal{V}$ is the set of nodes in the k -th subgraph, and $\bigcup_{k=1}^K \mathcal{S}_k = \mathcal{V}$, $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset$ for $i \neq j$.

Let $\mathbf{A} \in \mathbb{R}^{N \times N}$ be the adjacency matrix of the original graph, where $\mathbf{A}_{ij} = 1$ if $(v_i, v_j) \in \mathcal{E}$; otherwise, $\mathbf{A}_{ij} = 0$.

Let $\mathbf{A}_k \in \mathbb{R}^{N \times N}$ be the adjacency matrix of the k -th subgraph, where $\mathbf{A}_{k,ij} = 1$ if $(v_i, v_j) \in \mathcal{E}$ and $v_i, v_j \in \mathcal{S}_k$; otherwise, $\mathbf{A}_{k,ij} = 0$.

Let $\mathbf{H}^{(l)} \in \mathbb{R}^{N \times F_l}$ denote the node embedding matrix of the l -th layer of AAMPGCN, where F_l is the dimension of node embeddings in the l -th layer. Specifically, $\mathbf{H}^{(0)} = \mathbf{X}$ (the initial embedding is the node attribute matrix).

Let $\mathbf{W}^{(l)} \in \mathbb{R}^{F_{l-1} \times F_l}$ be the trainable weight matrix of the l -th layer, and $\sigma(\cdot)$ be the activation function (e.g., ReLU, Sigmoid).

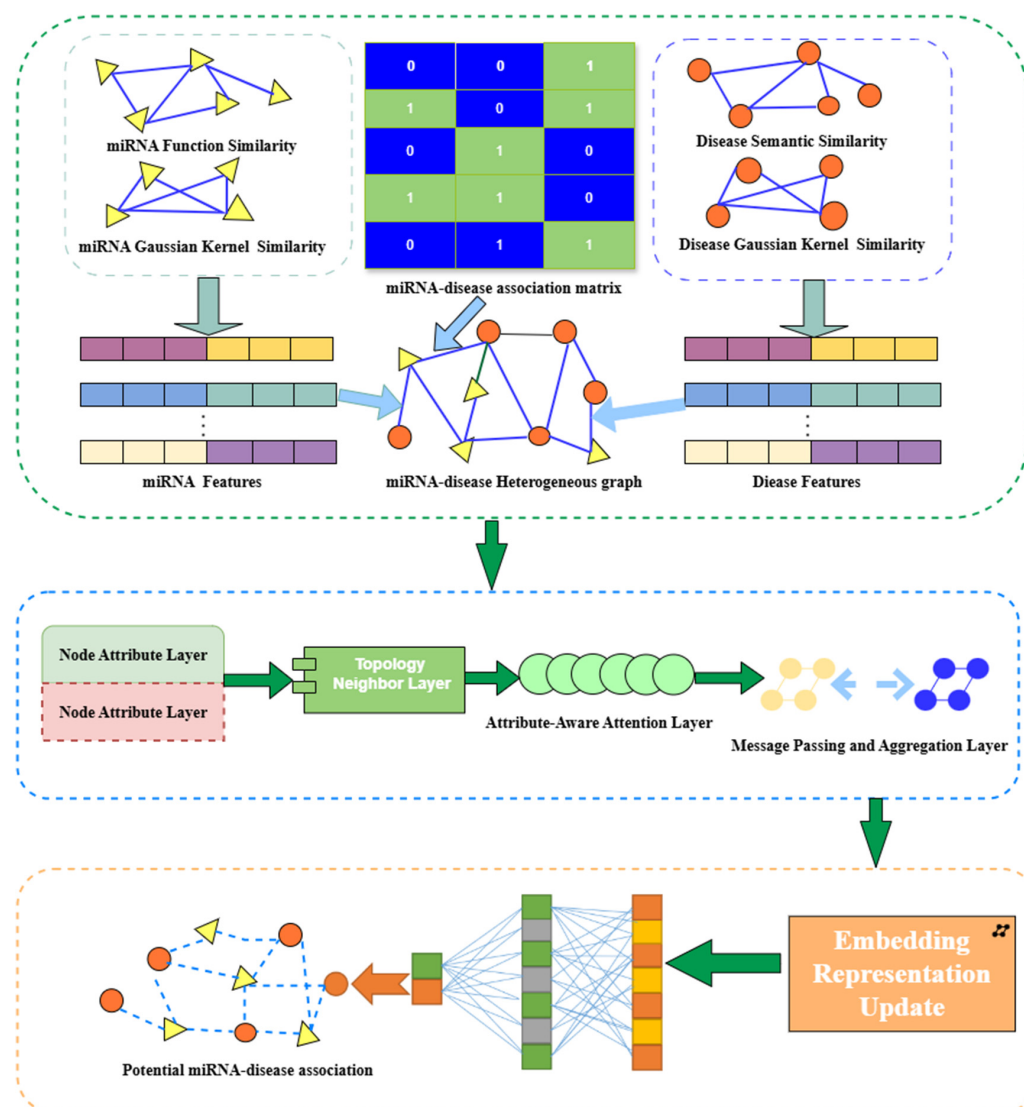


Figure 8. Flowchart of AAMPGCN model for predicting MDAs. Figure 8 illustrates the complete technical pipeline of our proposed AAMPGCN for miRNA-disease association prediction. The detailed descriptions are as follows: The green box corresponds to Step 1: Heterogeneous Graph Construction. We construct a bipartite miRNA-disease graph and derive multi-view node attribute features from miRNA functional similarity, disease semantic similarity and Gaussian kernel similarity. The blue box corresponds to Step 2: Attribute-Aware Message-Passing Convolution. Attribute-guided message propagation filters irrelevant neighbor signals adaptively according to biological feature consistency. Multi-layer graph convolution aggregates distinctive node representations while suppressing over-smoothing. The orange box corresponds to Step 3: Association score prediction. The aggregated miRNA and disease embeddings are combined to compute matching scores, which represent the likelihood of unvalidated miRNA-disease pairs.

The core of attribute-aware subgraph partitioning is to group nodes with similar attribute characteristics into the same subgraph. AAMPGCN first maps the original node attributes to a low-dimensional feature space through an attribute encoder to enhance the discriminability of attribute features. The attribute encoder is defined as a two-layer fully connected network:

$$\mathbf{Z} = \sigma_2(\mathbf{W}_2 \cdot \sigma_1(\mathbf{W}_1 \cdot \mathbf{X} + \mathbf{b}_1) + \mathbf{b}_2) \quad (14)$$

where $\mathbf{Z} \in \mathbb{R}^{N \times F_a}$ is the low-dimensional attribute embedding matrix (with F_a being the dimension of attribute embeddings), $\mathbf{W}_1 \in \mathbb{R}^{D \times F_a}$ and $\mathbf{W}_2 \in \mathbb{R}^{F_a \times F_a}$ are the weight matrices

of the attribute encoder, $\mathbf{b}_1$ and $\mathbf{b}_2$ are the bias terms, and σ_1, σ_2 are activation functions (usually ReLU).

On the basis of attribute embeddings $\mathbf{Z}$, K-means clustering is used to divide nodes into K subgraphs. The clustering objective is to minimize the within-cluster sum of squares (WCSS):

$$\min_{S_1, \dots, S_K} \sum_{k=1}^K \sum_{v_i \in S_k} \| \mathbf{z}_i - \boldsymbol{\mu}_k \|_2^2 \quad (15)$$

where $\mathbf{z}_i$ is the attribute embedding of the node v_i , and $\boldsymbol{\mu}_k = \frac{1}{|S_k|} \sum_{v_i \in S_k} \mathbf{z}_i$ is the centroid of the k -th subgraph. After clustering, we obtain the subgraph partition result $\mathcal{S}$ and the corresponding subgraph adjacency matrices $\{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_K\}$.

4.6.2. Constrained Message Passing Within Subgraphs

Unlike traditional GCNs that perform global message passing on the entire graph, AAMPGCN restricts message propagation within each attribute-homogeneous subgraph. This ensures that the message aggregation process only occurs between nodes with similar attributes, thereby reducing cross-attribute noise and alleviating over-smoothing. The message passing formula for the l -th layer in the k -th subgraph is

$$\mathbf{H}_k^{(l)} = \sigma(\hat{\mathbf{A}}_k \cdot \mathbf{H}^{(l-1)} \cdot \mathbf{W}^{(l)}) \quad (16)$$

where $\mathbf{H}_k^{(l)} \in \mathbb{R}^{N \times F_l}$ is the node embedding matrix of the l -th layer in the k -th subgraph, and $\hat{\mathbf{A}}_k$ is the normalized adjacency matrix of the k -th subgraph. The normalization of the adjacency matrix follows the symmetric normalization method used in traditional GCNs, which avoids the problem of numerical instability caused by excessive node degrees:

$$\hat{\mathbf{A}}_k = \mathbf{D}_k^{-1/2} \cdot (\mathbf{A}_k + \mathbf{I}) \cdot \mathbf{D}_k^{-1/2} \quad (17)$$

where $\mathbf{I} \in \mathbb{R}^{N \times N}$ is the identity matrix (used to retain the self-information of nodes), and $\mathbf{D}_k \in \mathbb{R}^{N \times N}$ is the degree matrix of the k -th subgraph, with $D_{k,ii} = \sum_{j=1}^N \mathbf{A}_{k,ij} + 1$ (the “+1” is due to the addition of the identity matrix).

For nodes not belonging to the k -th subgraph ($v_i \notin S_k$), their embeddings in $\mathbf{H}_k^{(l)}$ are set to 0, indicating that no message propagation occurs for these nodes in the k -th subgraph.

4.6.3. Multi-Subgraph Embedding Fusion

After completing the high-order message passing within each subgraph, AAMPGCN fuses the node embeddings obtained from all subgraphs to generate the final node embeddings. To fully utilize the information of each subgraph, a weighted fusion strategy is adopted, where the weight of each subgraph is determined by the similarity between the node's attribute embedding and the subgraph's centroid. The fusion formula is

$$\mathbf{H}^{(l)} = \sum_{k=1}^K \alpha_{k,i} \cdot \mathbf{H}_{k,i}^{(l)} \quad (18)$$

where $\mathbf{H}_{k,i}^{(l)}$ is the embedding of the node v_i in the k -th subgraph of the l -th layer, and $\alpha_{k,i}$ is the fusion weight of the k -th subgraph for the node v_i . The weight $\alpha_{k,i}$ is calculated using the softmax function to ensure that the sum of weights for each node is 1:

$$\alpha_{k,i} = \frac{\exp(-\| \mathbf{z}_i - \boldsymbol{\mu}_k \|_2^2)}{\sum_{m=1}^K \exp(-\| \mathbf{z}_i - \boldsymbol{\mu}_m \|_2^2)} \quad (19)$$

The weight calculation reflects that nodes with higher similarity to the centroid of a subgraph will have a higher weight for that subgraph's embedding, which ensures that the fused embedding can fully retain the attribute characteristics of the node.

4.6.4. Final Output and Loss Function

After L layers of subgraph message passing and embedding fusion, the final node embedding matrix $\mathbf{H}^{(L)}$ is obtained. For node classification or link prediction tasks (such as miRNA–disease association prediction), a predictor (usually a fully connected network) is used to map the final embeddings to the task output space:

$$\hat{\mathbf{Y}} = \text{softmax}(\mathbf{W}_{\text{pred}} \cdot \mathbf{H}^{(L)} + \mathbf{b}_{\text{pred}}) \quad (20)$$

where $\hat{\mathbf{Y}} \in \mathbb{R}^{N \times C}$ is the predicted output matrix (with C being the number of categories or the dimension of the link prediction output), $\mathbf{W}_{\text{pred}}$ and $\mathbf{b}_{\text{pred}}$ are the weight and bias of the predictor, respectively.

The loss function of AAMPGCN is designed according to the specific task. For classification tasks (such as node type classification), the cross-entropy loss is used:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C \mathbf{Y}_{i,c} \cdot \log(\hat{\mathbf{Y}}_{i,c}) \quad (21)$$

where $\mathbf{Y} \in \mathbb{R}^{N \times C}$ is the true label matrix. For link prediction tasks (such as miRNA–disease association prediction), the binary cross-entropy loss is used:

$$\mathcal{L} = -\frac{1}{M} \sum_{(i,j) \in \mathcal{E} \cup \mathcal{E}^-} (\mathbf{Y}_{ij} \cdot \log(\hat{\mathbf{Y}}_{ij}) + (1 - \mathbf{Y}_{ij}) \cdot \log(1 - \hat{\mathbf{Y}}_{ij})) \quad (22)$$

where M is the total number of positive and negative samples, $\mathcal{E}^-$ is the set of negative edges, $\mathbf{Y}_{ij} = 1$ if $(v_i, v_j) \in \mathcal{E}$ (positive sample), and $\mathbf{Y}_{ij} = 0$ if $(v_i, v_j) \in \mathcal{E}^-$ (negative sample).

4.6.5. Performance Evaluation

In order to evaluate the effectiveness of the proposed computational model, the following indices were adopted as the predictive performance, including accuracy (Acc), Recall (Rec) and Precision (Pre) and F1-measure ($F1$), which was defined as follows:

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \quad (23)$$

$$Rec = \frac{TP}{TP + FN} \quad (24)$$

$$Pre = \frac{TP}{FP + TP} \quad (25)$$

$$F1 = \frac{2 \times Pre \times Rec}{Pre + Rec} \quad (26)$$

The confusion matrix consists of four basic terms: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP denotes samples that are actually positive and predicted as positive; TN represents samples that are actually negative and predicted as negative; FP refers to samples that are actually negative but misclassified as positive; FN indicates samples that are actually positive but misclassified as negative. Based on the four metrics, Precision measures the proportion of correctly predicted positive samples among all predicted positive samples. Recall (Sensitivity) reflects the propor-

tion of correctly identified positive samples among all real positive samples. *F1-score* is the harmonic mean of Precision and Recall, comprehensively evaluating the overall classification performance.

5. Conclusions

Accurate prediction of miRNA–disease associations (MDAs) is crucial for deciphering the molecular mechanisms of complex diseases and identifying potential diagnostic biomarkers and therapeutic targets. This study proposed the AAMPGCN model for predicting miRNA–disease associations, aiming to improve the accuracy and reliability of MDA prediction. The superior performance of AAMPGCN is attributed to its core design features: integrating miRNA and disease attribute information into the message-passing process, partitioning attribute-homogeneous subgraphs, and restricting high-order message propagation within subgraphs, which effectively filters cross-attribute noise, avoids over-smoothing, and maintains the discriminability of node embeddings. Moreover, the AAMPGCN prioritizes miRNA and disease attributes, aggregating messages specifically among nodes with similar attribute characteristics that are relevant to miRNA–disease interactions. Case studies further confirmed that the predicted MDAs are biologically meaningful, with all top predicted associations supported by authoritative databases or published experimental literature, verifying the model’s practical application value. In conclusion, AAMPGCN provides a reliable, efficient computational tool for MDA prediction, offering important support for in-depth research on disease pathogenesis, biomarker discovery, and clinical precision medicine. It not only enriches the methodological system of MDA prediction but also lays a solid foundation for subsequent experimental verification and clinical translation of miRNA-related research.

Author Contributions: P.Q. and J.A. conceived the algorithm, carried out analyses, prepared the data sets, carried out experiments, and wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the High-level Talent Research Start-up Project of Liupanshui Normal University (Grant No. LPSSYKYJJ202524) and the Key Natural Science Project of Liupanshui Normal University (Grant No. LPSSYZDZK202204).

Institutional Review Board Statement: This study only used publicly available datasets and did not involve human participants or animal experiments, so ethical approval was not required.

Informed Consent Statement: This study did not involve human participants, and informed consent was not required.

Data Availability Statement: The datasets that support the findings of this study are available at <https://xzcpumt.cn:8888/downloaddata.html> (accessed on 10 June 2026).

Acknowledgments: The authors would like to thank all the guest editors and anonymous reviewers for their constructive advice.

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

Abbreviations

GCN	Graph Convolutional Network
AAMPGCN	Associations use Attribute-Aware Message Passing GCN
AUC	Area Under the Curve
AUPR	Area Under the Precision–Recall Curve
MDAs	MiRNA–Disease Associations
ROC	Receiver Operating Characteristic

PR	Precision–Recall
FS	Functional Similarity
MISIM	miRNA Similarity
DAG	Directed Acyclic Graph
GIP	Gaussian Interaction Profile

References

- Wang, J.; Samuels, D.C.; Zhao, S.; Xiang, Y.; Zhao, Y.-Y.; Guo, Y. Current Research on Non-Coding Ribonucleic Acid (RNA). *Genes* **2017**, *8*, 366. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nurk, S.; Koren, S.; Rhie, A.; Rautiainen, M.; Bizkadze, A.V.; Mikheenko, A.; Vollger, M.R.; Altemose, N.; Uralsky, L.; Gershman, A. The complete sequence of a human genome. *Science* **2022**, *376*, 44–53. [\[CrossRef\]](#) [\[PubMed\]](#)
- Krol, J.; Loedige, I.; Filipowicz, W. The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.* **2010**, *11*, 597–610. [\[CrossRef\]](#) [\[PubMed\]](#)
- Selbach, M.; Schwanhäusser, B.R.; Thierfelder, N.; Fang, Z.; Khanin, R.; Rajewsky, N. Widespread changes in protein synthesis induced by microRNAs. *Nature* **2008**, *455*, 58–63. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bartel, D.P. MicroRNAs: Target recognition and regulatory functions. *Cell* **2009**, *136*, 215–233. [\[CrossRef\]](#) [\[PubMed\]](#)
- Miska, E.A. How microRNAs control cell division, differentiation and death. *Curr. Opin. Genet. Dev.* **2005**, *15*, 563–568. [\[CrossRef\]](#) [\[PubMed\]](#)
- Karp, X.; Ambros, V. Developmental biology. Encount. microRNAs Cell Fate Signaling. *Science* **2005**, *310*, 1288–1289. [\[PubMed\]](#)
- Leva, G.D.; Croce, C.M. The Role of microRNAs in Cancer. In *Targeted Therapy in Translational Cancer Research*; Wiley-Blackwell: Hoboken, NJ, USA, 2015.
- Liu, B.; Yao, S.; Zhou, J. Micro-RNA 122 and micro-RNA 96 affected human osteosarcoma biological behavior and associated with prognosis of patients with osteosarcoma. *Biosci. Rep.* **2020**, *40*, BSR20201529. [\[PubMed\]](#)
- Zhao, J.; Chen, Y.; Liu, F.; Yin, M. Overexpression of miRNA-143 Inhibits Colon Cancer Cell Proliferation by Inhibiting Glucose Uptake. *Arch. Med. Res.* **2018**, *49*, 497–503. [\[CrossRef\]](#) [\[PubMed\]](#)
- Yue, J.; Zhang, B.; Wang, H.; Hou, X.; Wen, S. Dysregulated plasma levels of miRNA-132 and miRNA-134 in patients with obsessive-compulsive disorder. *Ann. Transl. Med.* **2020**, *8*, 996. [\[PubMed\]](#)
- Chen, H.; Zhang, Z. Similarity-based methods for potential human microRNA-disease association prediction. *BMC Med. Genom.* **2013**, *6*, 12.
- Li, H.; Guo, Y.; Cai, M.; Li, L. MicroRNA-disease association prediction by matrix tri-factorization. *BMC Genom.* **2020**, *21*, 617.
- Wang, Y.; Ren, Y.F.; Nie, J.T.; Li, K.Q. Enhancing Recommendation Performance Using Attribute-Aware Message-Passing and Augmentation GCN. *IEEE Trans. Netw. Sci. Eng.* **2026**, *13*, 2077–2092.
- Bustin, S.A.; Nolan, T. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J. Biomol. Tech.* **2004**, *15*, 155–166. [\[PubMed\]](#)
- Chen, Y.; Gelfond, J.A.; Mcmanus, L.M.; Shireman, P.K. Reproducibility of quantitative RT-PCR array in miRNA expression profiling and comparison with microarray analysis. *BMC Genom.* **2009**, *10*, 407.
- Várallyay, E.; Burgián, J.; Havelda, Z. MicroRNA detection by northern blotting using locked nucleic acid probes. *Nat. Protoc.* **2008**, *3*, 190–196. [\[CrossRef\]](#) [\[PubMed\]](#)
- You, Z.; Huang, Z.A.; Zhu, Z.X.; Yan, G.Y.; Chen, X. PBMDA: A novel and effective path-based computational model for miRNA-disease association prediction. *PLoS Comput. Biol.* **2017**, *13*, e1005455. [\[PubMed\]](#)
- Chen, X.; Jiang, Z.C.; Xie, D.; Huang, D.S.; Zhao, Q.; Yan, G.Y.; You, Z.H. A Novel computational model based on super-disease and miRNA for potential miRNA-disease association prediction. *Mol. Biosyst.* **2017**, *13*, 1202–1212. [\[CrossRef\]](#)
- Chen, X.; Gong, Y.; Zhang, D.H.; You, Z.H.; Li, Z.W. DRMDA: Deep representations-based miRNA-disease association prediction. *J. Cell. Mol. Med.* **2018**, *22*, 472–485. [\[PubMed\]](#)
- Wang, D.; Wang, J.A.; Lu, M.; Song, F.; Cui, Q.H. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics* **2010**, *26*, 1644–1650. [\[PubMed\]](#)
- Chen, X.; Yan, C.C.; Zhang, X.; You, Z.H.; Deng, L.X.; Liu, Y.; Zhang, Y.D.; Dai, Q.H. WBSMDA: Within and Between Score for MiRNA-Disease Association prediction. *Sci. Rep.* **2016**, *6*, 21106. [\[PubMed\]](#)
- Le, D.H. Integrating multiple microRNA functional similarity networks for improved disease–microRNA association prediction. *Biol. Methods Protoc.* **2025**, *10*, bpaf065. [\[PubMed\]](#)
- Chen, X.; Wang, L.; Qu, J.; Guan, N.N.; Li, J.Q. Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics* **2018**, *34*, 4256–4265. [\[PubMed\]](#)
- Xu, J.L.; Zhu, W.; Cai, L.J.; Liao, B.; Meng, Y.J.; Xiang, J.; Yuan, D.W.; Tian, G.; Yang, J.L. LRMCMDA: Predicting miRNA-Disease Association by Integrating Low-Rank Matrix Completion with miRNA and Disease Similarity Information. *IEEE Access* **2020**, *8*, 80728–80738.

26. Li, B.E.; Zhong, K.Y.; Deveci, M.; Tang, Y. Identifying potential miRNA-disease associations through an accurate matrix completion approach. *Brief. Bioinform.* **2025**, *26*, bbaf444. [[PubMed](#)]
27. Qin, C.; Zhang, J.C.; Ma, L.Y. EMCMDA: Predicting miRNA-disease associations via efficient matrix completion. *Sci. Rep.* **2024**, *14*, 12761. [[PubMed](#)]
28. Wei, L.; Hui, L.; Li, H.; Li, P.; Ting, T.; Qi, Z.; Li, Y. Identification of miRNA-disease associations via deep forest ensemble learning based on autoencoder. *Brief. Bioinform.* **2022**, *3*, bbac104.
29. Chen, X.; Huang, L.; Xie, D.; Zhao, Q. EGBMMDA: Extreme Gradient Boosting Machine for MiRNA-Disease Association prediction. *Cell Death Dis.* **2018**, *9*, 3. [[PubMed](#)]
30. Qiguo, D.; Zhaowei, W.; Ziqiang, L.; Xiaodong, D.; Jinmiao, S.; Maozu, G. Predicting miRNA-disease associations using an ensemble learning framework with resampling method. *Brief. Bioinform.* **2021**, *23*, bbab543.
31. Zheng, K.; You, Z.H.; Wang, L.; Zhou, Y.; Li, L.P.; Li, Z.W. MLMMDA: A machine learning approach to predict and validate MicroRNA-disease associations by integrating of heterogenous information sources. *J. Transl. Med.* **2019**, *17*, 260. [[PubMed](#)]
32. Wang, S.; Wang, F.; Qiao, S.; Zhuang, Y.; Zhang, K.; Pang, S.; Nowak, R.; Lv, Z. MSHGANMDA: Meta-Subgraphs Heterogeneous Graph Attention Network for miRNA-Disease Association Prediction. *J. Biomed. Health Inform.* **2023**, *27*, 4639–4648.
33. Han, H.; Zhu, R.; Liu, J.X.; Dai, L.Y. Predicting miRNA-disease associations via layer attention graph convolutional network model. *BMC Med. Inform. Decis. Mak.* **2022**, *22*, 6–9.
34. Gao, S.; Kuang, Z.; Duan, T.; Deng, L. DEJKMDR: miRNA-disease association prediction method based on graph convolutional network. *Front. Med.* **2023**, *10*, 1234050.
35. Liang, X.; Guo, M.; Jiang, L.; Fu, Y.; Zhang, P.; Chen, Y. Predicting miRNA-Disease Associations by Combining Graph and Hypergraph Convolutional Network. *Interdiscip. Sci. Comput. Life Sci.* **2024**, *16*, 289–303.
36. Xuan, P.; Han, K.; Guo, M.; Guo, Y.; Huang, Y. Prediction of microRNAs Associated with Human Diseases Based on Weighted k Most Similar Neighbors. *PLoS ONE* **2013**, *8*, e70204. [[CrossRef](#)] [[PubMed](#)]
37. He, Q.; Qiao, W.; Fang, H.; Bao, Y. Improving the identification of miRNA-disease associations with multi-task learning on gene-disease networks. *Brief. Bioinform.* **2023**, *24*, bbad203. [[PubMed](#)]
38. Ding, Y.; Tian, L.P.; Lei, X.; Liao, B.; Wu, F.X. Variational graph auto-encoders for miRNA-disease association prediction. *Methods* **2021**, *192*, 25–34. [[PubMed](#)]
39. Zhou, S.; Wang, S.; Wu, Q.; Azim, R.; Li, W. Predicting potential miRNA-disease associations by combining gradient boosting decision tree with logistic regression. *Comput. Biol. Chem.* **2020**, *85*, 107200. [[PubMed](#)]
40. Yang, Z.; Ren, F.; Liu, C.; He, S.; Sun, G.; Gao, Q.; Yao, L.; Zhang, Y.; Miao, R.; Cao, Y. dbDEMC: A database of differentially expressed miRNAs in human cancers. *BMC Genom.* **2010**, *11*, S5.
41. Ruepp, A.; Kowarsch, A.; Theis, F. PhenomiR: MicroRNAs in Human Diseases and Biological Processes. *Methods Mol. Biol.* **2012**, *822*, 249–260. [[PubMed](#)]
42. Mansoori, B.; Mohammadi, A.; Ghasabi, M.; Shirjang, S.; Dehghan, R. miR-142-3p as tumor suppressor miRNA in the regulation of tumorigenicity, invasion and migration of human breast cancer by targeting Bach-1 expression. *J. Cell. Physiol.* **2018**, *234*, 9816–9825. [[PubMed](#)]
43. Xiao, W.; Zhong, Y.; Wu, L.; Yang, D.; Ye, S.; Zhang, M. Prognostic value of microRNAs in lung cancer: A systematic review and meta-analysis. *Mol. Clin. Oncol.* **2019**, *10*, 67–77. [[PubMed](#)]
44. Zhou, L.H.; Li, Z.W.; Nie, R.; Wang, L.; You, Z.H. Predicting MiRNA-Disease Associations Using Chebyshev Graph Convolution and Graph. In *Proceedings of the 21st International Conference on Intelligent Computing-ICIC-Annual, Ningbo, China, 26–29 July 2025*; Springer Nature: Cham, Switzerland, 2025; pp. 24–35.
45. Yang, L.; Chengxiang, Q.; Jian, T.; Bin, G.; Jichun, Y.; Tianzi, J.; Qinghua, C. HMDD v2.0: A database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res.* **2014**, *42*, D1070–D1074.
46. Chen, X.; Yan, C.C.; Zhang, X.; You, Z.H. Long non-coding RNAs and complex diseases: From experimental results to computational models. *Brief. Bioinform.* **2016**, *18*, 558–576.
47. Chen, X.; Yan, C.C.; Luo, C.; Ji, W.; Zhang, Y.; Dai, Q. Constructing lncRNA functional similarity network based on lncRNA-disease associations and disease semantic similarity. *Sci. Rep.* **2015**, *5*, 11338. [[PubMed](#)]
48. Chen, X.; Huang, Y.A.; Wang, X.S.; You, Z.H.; Chan, K.C. FMLNCSIM: Fuzzy measure-based lncRNA functional similarity calculation model. *Oncotarget* **2016**, *7*, 45948–45958. [[PubMed](#)]
49. Chen, M.; Peng, Y.; Li, A.; Li, Z.; Deng, Y.; Liu, W.; Liao, B.; Dai, C. A novel information diffusion method based on network consistency for identifying disease related microRNAs. *RSC Adv.* **2018**, *8*, 36675–36690. [[CrossRef](#)] [[PubMed](#)]
50. Van Laarhoven, T.; Nabuurs, S.B.; Marchiori, E. Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics* **2011**, *27*, 3036–3043. [[PubMed](#)]
51. Chen, X.; Huang, Y.A.; You, Z.H.; Yan, G.Y.; Wang, X.S. A novel approach based on KATZ measure to predict associations of human microbiota with non-infectious diseases. *Bioinformatics* **2016**, *33*, 733–739. [[CrossRef](#)]

52. Chen, X.; Yan, G.Y. Novel human lncRNA-disease association inference based on lncRNA expression profiles. *Bioinformatics* **2013**, *29*, 2617–2624. [[PubMed](#)]
53. Zhou, L.W.; Tang, Y.C.; Yan, G.Y. A New Estimation Method for the Biological Interaction Predicting Problems. *IEEE-ACM Trans. Comput. Biol. Bioinform.* **2022**, *19*, 1415–1423. [[PubMed](#)]

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.