Dual-Channel Learning Framework for Zero-Shot CircRNA-miRNA Interaction Prediction via State Space Modeling

Mengmeng Wei[1*], Lei Wang[1*†], Zhu-Hong You[2‡], Pengwei Hu[3], Bowei Zhao[3], Zhi-An Huang[4], Yu-An Huang[2], Haicheng Yi[2]

[1]School of Computer Science and Technology/School of Artificial Intelligence, China University of Mining and Technology, Xuzhou, China
[2]School of Computer Science, Northwestern Polytechnical University, Xi'an, China
[3]Xinjiang Technical Institute of Physics and Chemistry, University of the Chinese Academy of Sciences, Urumqi, China
[4]Research Office, City University of Hong Kong (Dongguan), Dongguan, China
leiwang@cumt.edu.cn, zhuhongyou@nwpu.edu.cn

Abstract

CircRNA-miRNA interaction (CMI) plays a pivotal role in disease therapeutics and drug discovery. However, existing methods face several challenges in modeling complex biological networks and zero-shot learning scenarios. Biological networks encapsulate rich biological information, yet current approaches often fail to fully exploit this depth. Moreover, zero-shot prediction requires models to identify new interactions without relying on previously observed samples, imposing stringent requirements on generalization capabilities. To address these limitations, we propose a dual-channel learning framework leveraging ***St****ate space* ***m****odeling* for ***Zero****-shot* CMI prediction (ZeroStem). ZeroStem first enhances the biological relevance of node using prior knowledge, and employs a graph Transformer to extract macro-topological representations. Subsequently, it generates semantic subgraphs based on meta-paths to focus on specific biological relationships, utilizing the Mamba to extract micro-semantic representations via state space modeling. Finally, macro-topological and micro-semantic representations are seamlessly integrated through linear transformation and residual connections, enabling high-precision zero-shot CMI prediction. Extensive experiments on multiple benchmark datasets demonstrate that ZeroStem significantly outperforms existing methods, validating its efficiency and robust generalization in CMI prediction. Case studies further illustrate that ZeroStem offers novel insights into the molecular mechanisms underlying intricate disease-associated networks.

Introduction

Circular RNAs (circRNAs) and microRNAs (miRNAs) are abundant non-coding RNAs critical for post-transcriptional gene regulation (Nemeth et al. 2024). They are implicated in a wide range of human diseases, including cancer, neurological disorders, and cardiovascular pathologies (Caporali et al. 2024). In particular, circRNA-miRNA interactions (CMIs) serve as key modulators in competitive endogenous RNA (ceRNA) networks by sequestering miRNAs, thereby regulating gene expression and influencing disease progression (Nejadi Orang and Abdoli Shadbad 2024). Accurate prediction of CMIs is therefore indispensable for understanding disease mechanisms and guiding therapeutic target discovery and drug design.

In recent years, numerous computational methods have been proposed for CMI prediction, such as utilizing matrix computations to extract latent interaction information (Qian et al. 2022), leveraging graph neural networks to capture topological structures from association networks (Wang et al. 2024a), and employing natural language processing (NLP) techniques to analyze biological characteristics embedded in molecular sequences (Wei et al. 2025). Despite the promising advancements, existing methods fall short in fully harnessing the intricate patterns and underlying semantics embedded within biomolecular networks. The complexity of biological systems stems not only from direct molecular interactions but also from the higher-order topological structures that span across different molecular layers and the context-specific subnetworks that reflect localized biological functions (Su et al. 2025). Traditional approaches often focus on isolated network components, struggling to integrate global connectivity patterns with fine-grained semantic relationships.

Furthermore, CMI studies typically neglect the constraints imposed by shared disease contexts on these interactions, focusing primarily on individual molecular relationships. To address this gap, we leverage the circRNA-disease association and miRNA-disease association data to construct a novel dataset for zero-shot CMI prediction, aiming to predict unknown regulatory relationships. The dataset supports a comprehensive analysis of molecular interactions within disease frameworks, while requiring computational models to capture both broad network structures and disease-specific semantic relationships.

In this paper, we propose a novel dual-channel learning framework based on ***St****ate space* ***m****odeling*, designed to pre-

[*]These authors contributed equally.
[†]Corresponding Author.
[‡]Corresponding Author.

dict CMIs in ***Zero**-shot* scenarios (ZeroStem). Specifically, ZeroStem initially incorporates prior knowledge into molecular representations, thereby enhancing the biological relevance of node. It then extracts global macro-topological features using a graph Transformer, which effectively captures long-range dependencies across the biomolecular network. To further characterize localized regulatory patterns, we construct semantically guided subgraphs based on meta-paths and leverage Mamba to perform state space modeling, thereby yielding micro-semantic features. Finally, ZeroStem seamlessly integrates these macro-topological and micro-semantic features through linear transformation and residual connections, employing a multi-layer perceptron (MLP) to enable CMI prediction. The main contributions of this paper are described as follows:

- We propose ZeroStem, a novel dual-channel learning framework constructed upon state space modeling, to accurately predict CMI under zero-shot scenarios. we also construct a new dataset designed to support this task, CMI-Zero.
- We introduce a representation learning approach that leverages biological prior knowledge, graph structural dependencies, and meta-path guided micro-semantic patterns to enable a robust fusion of multi-source features.
- Extensive experimental results show ZeroStem outperforms baseline models across benchmark datasets, highlighting its effectiveness in discovering unknown CMIs and suggesting potential for elucidating disease mechanisms and identifying new biomarkers.

Related Work

Graph Neural Network-Based Methods

Graph neural network-based methods model the intricate topological relationships within heterogeneous graphs, providing notable advantages for CMI prediction. For example, MRHRL adopts multi-relational hypergraphs to capture high-order molecular associations, utilizing view attention mechanisms to fuse complementary information, thereby enhancing the representation of complex network (Yin et al. 2024). Similarly, LSAGHI introduces an autonomous adjustment mechanism to dynamically optimize hypergraph structures and node associations, improving the adaptability of high-order relationship representations (Yin et al. 2025).

Biomolecular Sequence Modeling-Based Methods

Biomolecular sequence modeling-based methods treat circRNA and miRNA sequences as genomic languages, applying NLP techniques to uncover intrinsic characteristics, thereby establishing a transferable semantic representation for CMI prediction. BioKG-CMI integrates biomedical knowledge graphs, employing BERT to learn generalized sequence representations and utilizing the DisMult to mine biological logical rules, effectively advancing the capability for CMI prediction (Wei et al. 2024). DGCLCMI adopts word2vec for sequence encoding, coupled with cross-domain collaborative filtering to optimize feature extraction, enabling accurate CMI prediction (Cao et al. 2025).

Zero-Shot Learning Methods

Zero-shot learning methods remain underexplored in the CMI domain but have shown promise in adjacent bioinformatics tasks. ZeroDDI proposes a semantic enhancement module with a dual-modal alignment strategy to mitigate class imbalance and enhance prediction of zero-shot drug-drug interaction events (Wang et al. 2024b). DSHGZS constructs dual-stream heterogeneous graphs based on zero-shot embeddings derived from large language models to generate high-quality miRNA and drug representation, addressing miRNA-drug sensitivity prediction (Peng et al. 2024).

Methodology

The architecture of ZeroStem, as illustrated in Figure 1, captures macro-topological structures through attention mechanisms and generates micro-semantic representations via state space modeling on meta-path-driven semantic subgraphs, thereby enabling precise prediction of potential CMIs in zero-shot scenarios.

Preliminaries

In this study, we adopt diseases as intermediary anchors to learn representations of circRNAs and miRNAs under zero-shot conditions. To this end, we systematically process biomolecular prior knowledge, enriching node embeddings with biological relevance.

Disease Semantic Representation. We leverage the Disease Ontology (DO) similarity to describe disease nodes (Lu, Han, and Ning 2022). For a specific disease d, its directed acyclic graph (DAG) is expressed as $\text{DAG}(d) = (d, N_d, E_d)$. Here, N_d encompasses d along with all its ancestor nodes, and E_d consists of the semantic relationship edges linking these nodes. The contribution of a disease a to disease d, denoted $C_d(a)$, is computed recursively as:

$$C_d(a) = \begin{cases} 1 & a = d \\ \max_{a' \in \text{Children}(a)} \mu \cdot C_d(a') & a \neq d \end{cases} \quad (1)$$

where $\mu \in (0, 1)$ represents a semantic decay factor, optimized to 0.5 through grid search, and $\text{Children}(a)$ refers to the direct child nodes of node a. The DO similarity score S_d is defined as the cumulative contribution of all nodes within its associated DAG structure, calculated as:

$$S_d = \sum_{a \in N_d} C_d(a) \quad (2)$$

Given that diseases at the same hierarchical level in the DAG possess equivalent contribution values, the semantic similarity between diseases d_i and d_j is calculated as:

$$Sim_{\text{DO}}(d_i, d_j) = \frac{\sum_{a \in N_{d_i} \cap N_{d_j}} [C_{d_i}(a) + C_{d_j}(a)]}{S_{d_i} + S_{d_j}} \quad (3)$$

CircRNA Functional Similarity. Based on the assumption that functionally related circRNAs are often associated with phenotypically similar diseases, we calculate the functional similarity between circRNAs using a Gaussian interaction profile (GIP) kernel (Wang et al. 2020). Specifically, for each

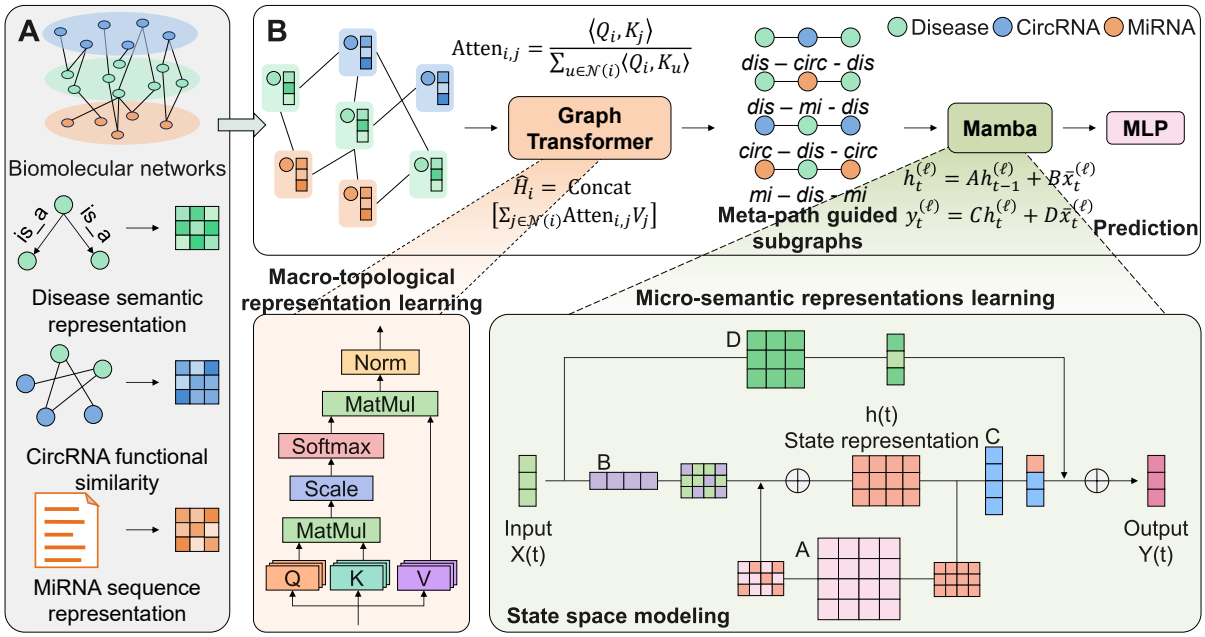

Figure 1: The overall framework of ZeroStem.

circRNA c_i, its association with all diseases is represented as a binary interaction profile vector $P(c_i) \in \{0,1\}^m$, where m denotes the total number of diseases. If circRNA c_i is known to be associated with a disease d, then $P(c_i, d) = 1$; otherwise, it is set to 0. The functional similarity between two circRNAs c_i and c_j is then computed using the radial basis function (RBF) as follows:

$$\mathrm{GKS}\left(c_i, c_j\right) = \exp\left(-\gamma \left\|P\left(c_i\right) - P\left(c_j\right)\right\|^2\right) \tag{4}$$

where $\|\cdot\|$ denotes the Euclidean distance, and γ is the kernel width parameter that controls the decay rate of similarity, computed as follows:

$$\gamma = \frac{1}{n}\sum_{i=1}^{n}\left\|P\left(c_i\right)\right\|^2 \tag{5}$$

MiRNA Sequence Representation. To comprehensively explore the biological information embedded in miRNA sequences, we utilize a generator-discriminator framework from ELECTRA as the semantic encoder (Clark et al. 2020). It consists of a generator G and a discriminator D, both implemented using Transformer encoder architectures. The input sequence $S = [s_1, s_2, \ldots, s_g]$ is mapped to a context-aware vector representation $h(s) = [h_1, h_2, \ldots, h_g]$. When the nucleotide at position τ, denoted as s_τ, is replaced with the special token `[MASK]`, the generator computes a probability distribution over the vocabulary via a softmax layer to predict the original nucleotide at that position, formulated as follows:

$$p_G\left(s_\tau \mid s\right) = \frac{\exp\left(e\left(s_\tau\right)^T h_G(s)_t\right)}{\sum_{s'}\exp\left(e\left(s'\right)^T h_G(s)_\tau\right)} \tag{6}$$

where $e(s_\tau)$ denotes the embedding representation of s_τ. The discriminator D evaluates the probability that each token generated by the generator G is authentic, computed as follows:

$$D(s,\tau) = \mathrm{Sigmoid}\left(w^T h_D(s)_\tau\right) \tag{7}$$

Dual-Channel Learning Framework

The dual-channel learning framework is engineered to fuse diverse information sources to enhance predictive performance. It leverages two distinct processing channels: the macro-topological channel, which captures the global structural dependencies within the biomolecular network, and the micro-semantic channel, which focuses on the finer, context-specific relationships derived from meta-path guided subgraphs. As illustrated in Algorithm 1, the specific details are presented in the following paragraphs.

Macro-topological Representation Learning. To capture the macro-topological structure of the biomolecular network composed of circRNAs, miRNAs, and diseases, we propose a graph learning framework based on the Transformer architecture (Shi et al. 2021). Specifically, the learning process begins with the construction of an association graph $G = (V, E)$, where V denotes the set of nodes and E denotes the set of edges. The initial node features $F = [f_1, f_2, \ldots, f_k]$ are derived from the representations introduced in the ***Preliminaries***. The graph Transformer employs a multi-head attention mechanism to compute attention scores, allowing the model to selectively aggregate information from the most relevant neighboring nodes. The attention score from node i to node j is computed as:

$$\begin{aligned} Q_i &= W_q f_i + b_q \\ K_j &= W_k f_j + b_k \end{aligned} \tag{8}$$

Algorithm 1: Dual-Channel Learning Framework

Input: Association graph $G = (V, E)$; node features $F \in \mathbb{R}^{|V| \times d}$.

Parameters: Meta-path $\mathcal{P} = \{p_1, \ldots, p_m\}$; epochs T; learning rate η; loss weights λ_1, λ_2.

Output: Learned node embeddings $\widehat{H} \in \mathbb{R}^{|V| \times d}$.

for epoch = 1 to T **do**
  $H^0 \leftarrow \text{Graph Transformer}(G, F)$
  for each meta-path $p \in \mathcal{P}$ **do**
    Extract subgraph $G_p = (V_p, E_p)$ induced by p
    $Z_p^0 \leftarrow H^0[V_p]$
    for $l = 1$ to L **do**
      $Z_p^l \leftarrow \text{Mamba}(Z_p^{l-1})$
      $Z_p^l \leftarrow \text{Dropout}(Z_p^l)$
    end for
    $\widetilde{H}_p \leftarrow \text{Scatter}(Z_p^L, V_p)$
  end for
  $\widetilde{H} \leftarrow \text{Concat}(\{\widetilde{H}_p\}_{p \in \mathcal{P}})$
  $\widehat{H} \leftarrow \text{Dropout}(\text{Linear}(\widetilde{H}) + H^0)$
  $\mathcal{L}_{\text{recon}} \leftarrow \text{MSE}(\widehat{H}, X)$
  $\mathcal{L}_{\text{reg}} \leftarrow \|\theta\|_1$
  $\mathcal{L} \leftarrow \lambda_1 \mathcal{L}_{\text{recon}} + \lambda_2 \mathcal{L}_{\text{reg}}$
end for
return $\widehat{H}$

$$\text{Atten}_{i,j} = \frac{\langle Q_i, K_j \rangle}{\sum_{u \in \mathcal{N}(i)} \langle Q_i, K_u \rangle} \tag{9}$$

where $\langle Q, K \rangle = \exp\left(\frac{Q^\top K}{\sqrt{d_{att}}}\right)$, and d_{att} denotes the dimensionality of each attention head. W and b are trainable weight matrices and bias terms, respectively.

The message passing and aggregation from node j to node i is computed as follows:

$$V_j = W_v f_j + b_v \tag{10}$$

$$\widehat{H}_i = \text{Concat}\left[\sum_{j \in \mathcal{N}(i)} \text{Atten}_{i,j} V_j\right] \tag{11}$$

Micro-semantic Representations Learning. Although the macro-topological module effectively extracts the long-range dependencies, it may fall short in modeling fine-grained semantic relationships. To address this limitation, we design a meta-path guided Mamba module, which leverages state space modeling to learn micro-semantic features rooted in localized, context-specific regulatory patterns (Gu and Dao 2023).

Let $\mathcal{P} = \{p_1, p_2, \ldots, p_m\}$ denote a set of meta-paths within a heterogeneous molecular network comprising diseases, circRNAs, and miRNAs. In this study, we define four types of meta-paths: d_c_d, d_m_d, c_d_c, and m_d_m. Each meta-path encodes a specific relational context among molecule entities, facilitating the discovery of regulatory or functional similarities guided by shared disease associations.

For the meta-path $p \in \mathcal{P}$, the corresponding semantic subgraph $G_p = (V_p, E_p)$ is defined such that V_p contains all nodes that participate in at least one instance of p, and E_p includes the edges that connect consecutive nodes along these instances.

Given the macro-topological embeddings $\widehat{H}$, the initial features of the subgraph G_p are set as $Z_p = \widehat{H}[V_p]$, where $\widehat{H}[V_p]$ denotes the embeddings associated with the nodes in V_p. These features are then processed through a stack of Mamba L layers to capture the propagation of localized sequential semantic information:

$$Z_p^{(\ell)} = \text{Mamba}(Z_p^{(\ell-1)}), \quad \ell = 1, 2, \ldots, L \tag{12}$$

The input features Z_p are projected into two parallel streams:

$$Z_p'^{(\ell-1)} = \text{Linear}\left(Z_p^{(\ell-1)}\right) = [x^{(\ell)}, z^{(\ell)}] \tag{13}$$

where $z^{(\ell)}$ serves as a gating signal to selectively control the retention of information in the state space model output. The $x^{(\ell)}$ is passed through a causal one-dimensional convolution to encode local context:

$$\bar{x}^{(\ell)} = \phi_{\text{causal}}(x^{(\ell)}) \tag{14}$$

The evolution of each channel is modeled through a learned linear dynamical system, where the hidden state $h_t^{(\ell)}$ at t step is updated based on the previous hidden state $h_{t-1}^{(\ell)}$ and the current input $\bar{x}_t^{(\ell)}$. This is expressed as:

$$h_t^{(\ell)} = A h_{t-1}^{(\ell)} + B \bar{x}_t^{(\ell)} \tag{15}$$

where A and B are learnable parameters of the hidden state. The matrix A captures the dependencies between successive hidden states, while B controls the influence of the current input on the hidden state. The output $y_t^{(\ell)}$ is then computed as:

$$y_t^{(\ell)} = C h_t^{(\ell)} + D \bar{x}_t^{(\ell)} \tag{16}$$

where C and D are learnable parameters that determine the contribution of the hidden state and the input to the output, respectively. The outputs from all channels are then projected to produce the final representation:

$$\tilde{Z}_p^{(\ell)} = \text{Linear}(y^{(\ell)}) \tag{17}$$

The $z^{(\ell)}$ controls the residual gating via a nonlinear function:

$$Z_p^{(\ell)} = \text{LayerNorm}\left(Z_p^{(\ell-1)} + \tilde{Z}_p^{(\ell)} \odot \sigma(z^{(\ell)})\right) \tag{18}$$

After passing through stacked L layers, a refined node representation $Z_p^{(L)}$ is yielded. To mitigate overfitting and enhance generalization, a dropout layer is applied after the final Mamba layer:

$$Z_p^{(L)} \leftarrow \text{Dropout}(Z_p^{(L)}, \rho) \tag{19}$$

where elements of the embedding are randomly set to zero with probability ρ, and remaining activations are rescaled by $1/(1-\rho)$ to preserve the expectation.

Since the subgraph G_p only covers a subset of the full graph G, the refined embeddings $Z_p^{(L)}$ must be projected

back into the global node space to enable fusion across all paths. This is achieved via a sparse-scatter operation:

$$\widetilde{H}_p[i] = \begin{cases} Z_p^{(L)}[j], & \text{if node } i = V_p[j] \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

Finally, the semantic embeddings from all subgraphs are integrated with the macro-topological features through a residual-enhanced projection, yielding the node-level representations $\widehat{H}$. This is computed as follows:

$$\widehat{H} = \text{Linear}\left(\widetilde{H}_p\right) + \widehat{H} \tag{21}$$

Through the state space modeling guided by meta-path induced subgraphs, our model captures the micro-level semantic nuances along biologically meaningful paths, synergistically enhancing the macro-topological representation provided by the graph Transformer.

Training Objective and Prediction

The loss function employed for training the proposed model consists of two primary components: the reconstruction loss and the L1 regularization loss. The reconstruction loss measures the discrepancy between the output embeddings and the original input features, ensuring that the learned representations retain essential information. It is defined using the Mean Squared Error (MSE) as follows:

$$\mathcal{L}_{\text{recon}} = \frac{1}{|\mathcal{V}_{\text{n}}|} \sum_{i \in \mathcal{V}_{\text{n}}} \left\| \widehat{\mathbf{H}}_i - \mathbf{x}_i \right\|_2^2, \tag{22}$$

where $\mathcal{V}_{\text{n}}$ is the set of training nodes. The reconstruction loss encourages the model to generate embeddings that closely approximate the input features, thereby preserving the structural and semantic information inherent in the data.

To enhance the generalization capability, an L1 regularization term is applied to promote sparsity in the parameter space. The L1 regularization drives a large number of parameters to become zero, thereby effectively performing feature selection and reducing the complexity. The total loss function is a weighted sum of the reconstruction loss and the regularization loss:

$$\mathcal{L} = w_{\text{recon}} \cdot \mathcal{L}_{\text{recon}} + w_{\text{reg}} \cdot \mathcal{L}_{\text{reg}} \tag{23}$$

where w_{recon} and w_{reg} is the loss weight assigned to the regularization loss and L1 regularization.

The $\widehat{H} \in \mathbb{R}^{|V| \times d}$ denote the final node embeddings obtained through dual-channel fusion. For any node pair (i, j), we define the concatenated representation as $[\,\widehat{H}_i \,\|\, \widehat{H}_j\,]$. To predict the association between node pairs (i, j), we employ a multi-layer perceptron (MLP) classifier to compute the interaction score.

Experiments

Datasets and Settings

Datasets. In this study, we introduce a novel dataset, CMI-Zero, to support zero-shot CMI prediction. It is constructed by integrating circRNA-disease association and miRNA-disease association data from multiple biological databases, including circAtlas 3.0 (Wu, Zhao, and Zhang 2024), circRNADisease 2.0 (Sun et al. 2024), and HMDD 4.0 (Cui et al. 2024). After data cleaning and integration, CMI-Zero includes 2,014 diseases, 442 circRNAs, and 310 miRNAs, with 1,812 circRNA-disease associations and 17,329 miRNA-disease associations. A total of 721 CMI pairs were identified and designated as positive samples. These samples are reserved exclusively for model evaluation, ensuring they are excluded from the training phase to provide an unbiased performance assessment.

In addition, CMI-9905 and CMI-9589 are utilized to further assess the generalization capability of the model. The CMI-9905 dataset integrates data from CircBank (Liu et al. 2019) and circR2Cancer (Lan et al. 2020), containing 2,346 circRNAs, 962 miRNAs, and 9,905 interaction pairs derived from literature, experimental validation, and CircBank. The CMI-9589 dataset, sourced from CircBank, includes 2,115 circRNAs, 821 miRNAs, and 9,589 high-confidence interaction pairs.

Evaluation Metrics. To comprehensively evaluate model performance, we employ accuracy (Acc.), F1-score (F1), the area under the receiver operating characteristic curve (AUC), and the area under the precision-recall curve (AUPR).

Experimental Settings. The evaluation is performed under a five-fold cross-validation to ensure the stability of the results. Negative samples are constructed using a biologically constrained framework based on subcellular localization, improving upon traditional random sampling approaches. ZeroStem is implemented in Python 3.10.18 with PyTorch 2.3.1, and all experiments are conducted using an NVIDIA RTX 5070 GPU.

Baselines. To assess the effectiveness of ZeroStem, we compare it with several representative baselines, categorized as follows:

- **Network Embedding-based:** We select two representative methods, SDNE (Wang, Cui, and Zhu 2016) and MUSAE (Rozemberczki, Allen, and Sarkar 2021), as baselines. SDNE uses a deep autoencoder to preserve both local and global nonlinear network structures. MUSAE captures multi-scale attribute neighborhood relationships through a skip-gram inspired approach, achieving efficient embeddings across various graphs.
- **Multi-source Feature Fusion-based:** Two well-established methods are BioKG-CMI (Wei et al. 2024) and EGATCMI (Wei et al. 2025). BioKG-CMI and EGATCMI fuse sequence features obtained from Transformer with knowledge graph embeddings and graph attention network-based embeddings, respectively, to generate multi-source feature representations.
- **Denoising and Adaptive Learning-based:** Representative methods are MetaGC (Jo, Bu, and Shin 2023) and ADA-GAD (He et al. 2024). MetaGC improves graph clustering by learning adaptive weights to reduce noise edges. ADA-GAD uses a two-stage denoising and multi-level pretraining framework to alleviate overfitting and enhance anomaly detection.

Method	CMI-Zero				CMI-9905				CMI-9589			
	Acc.	F1	AUC	AUPR	Acc.	F1	AUC	AUPR	Acc.	F1	AUC	AUPR
SDNE	0.8245	0.8219	0.9078	0.9165	0.8591	0.8343	0.8927	0.9018	0.8342	0.8297	0.9153	0.9238
MUSAE	***0.8502***	***0.8459***	0.9039	0.9154	***0.8601***	0.8370	0.9072	0.8923	0.8962	***0.8985***	0.9529	0.9629
BioKG-CMI	0.8301	0.8159	0.8935	0.9212	0.8219	0.7923	0.9063	0.9278	0.8561	0.8358	0.9463	0.9581
EGATCMI	0.8261	0.7266	0.8607	0.8745	0.8277	0.8397	0.9106	0.8967	0.8778	0.8798	0.9470	0.9426
MetaGC	0.8398	0.8275	0.9068	***0.9311***	0.8342	0.8292	0.9198	0.9274	***0.8966***	0.8955	***0.9641***	***0.9687***
ADA-GAD	0.8377	0.8360	***0.9092***	0.9253	0.8519	***0.8522***	***0.9356***	***0.9316***	0.8603	0.8515	0.9473	0.9495
ZeroStem	**0.9348**	**0.9318**	**0.9688**	**0.9788**	**0.9386**	**0.9371**	**0.9878**	**0.9893**	**0.9423**	**0.9429**	**0.9887**	**0.9899**
Improv.(%)	8.46	8.59	5.96	4.77	7.85	8.49	5.22	5.77	4.57	4.44	2.46	2.12

Table 1: The performance of ZeroStem and the baseline methods on the three datasets across five-fold cross-validation.

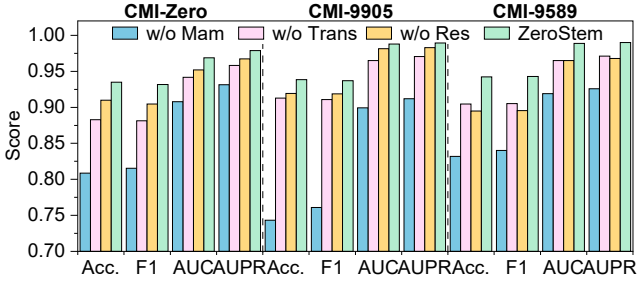

Figure 2: Performance comparison of different ablation variants on the three datasets using five-fold cross-validation.

Overall Performance

Table 1 presents the average results of ZeroStem and baseline methods on benchmark datasets under the five-fold cross-validation. **Bold** indicate the best results among all methods, while ***bold*** numbers represent the best results among the compared baseline methods. ZeroStem significantly outperforms the six baseline methods on CMI prediction across the CMI-Zero, CMI-9905, and CMI-9589 datasets. Specifically, on the CMI-Zero dataset, ZeroStem achieves an accuracy of 0.9348, F1 of 0.9318, AUC of 0.9688, and AUPR of 0.9788, with improvements of 8.46%, 8.59%, 5.96%, and 4.77% over the best baseline, respectively. On the CMI-9905 dataset, it obtains an accuracy of 0.9386, F1 of 0.9371, AUC of 0.9878, and AUPR of 0.9893, showing improvements of 7.85%, 8.49%, 5.22%, and 5.77%, respectively. On the CMI-9589 dataset, ZeroStem achieves an accuracy of 0.9423, F1 of 0.9429, AUC of 0.9887, and AUPR of 0.9899, with respective improvements of 4.57%, 4.44%, 2.46%, and 2.12%. These results demonstrate the efficiency and generalizability of ZeroStem under different data, highlighting its potential as a reliable method for CMI prediction.

Ablation Study

To investigate the role of each component within the ZeroStem framework, we conduct ablation experiments by separately removing the Mamba (w/o Mam), the graph Transformer (w/o Trans), and the residual connection (w/o Res). The comparison results are presented in Figure 2.

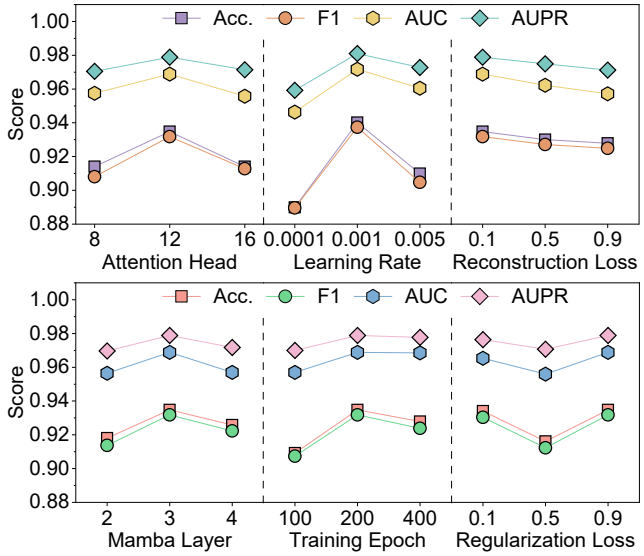

Figure 3: Parameter sensitivity analysis of ZeroStem on the CMI-Zero dataset across five-fold cross-validation.

As illustrated, each of these components contributes to performance improvements on the three datasets. The absence of the Mamba module leads to a marked decline in performance, highlighting the importance of capturing fine-grained semantic features. When the graph Transformer is removed, the ability to model global structural information is impaired, with a more pronounced performance drop on the CMI-Zero dataset, which involves more complex interaction patterns. Removing the residual connection affects the direct propagation of input features, resulting in a consistent, albeit moderate, reduction in performance. These observations demonstrate that the combination of these modules is essential for achieving stable and accurate predictions.

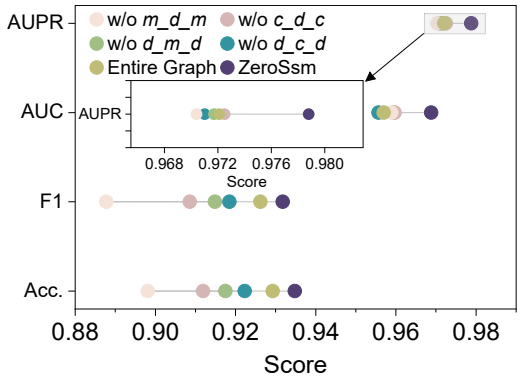

Figure 4: Performance comparison of ZeroStem with different meta-paths and the entire graph.

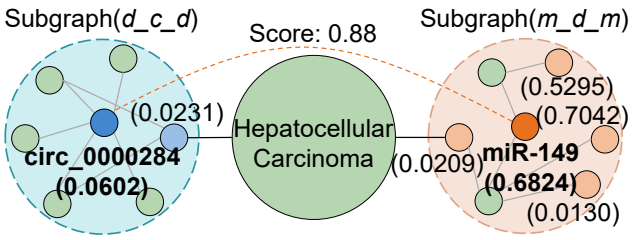

Figure 5: Visualization of molecular contributions in the m_d_m and d_c_d subgraphs for HCC using SHAP analysis.

Parameter Sensitivity Analysis

We perform a parameter analysis across three aspects: architecture parameters, training configurations, and loss weight ratios. All experiments are conducted on the CMI-Zero dataset using five-fold cross-validation, and the results are presented in Figure 3.

Architecture Parameters. For the architecture, setting the attention heads to 12 in the graph Transformer delivers the most reliable performance. Fewer heads, like 8, tend to miss out on relational diversity, while increasing to 16 introduces redundancy and noise. In the state space modeling-based Mamba module, the 3-layer Mamba offers sufficient semantic depth without introducing excessive model complexity.

Training Parameters. A learning rate of 0.001 ensures optimal performance and robustness. Lower rates slow convergence, causing underfitting, while higher rates lead to unstable updates and inconsistent performance. Training for 200 epochs achieves the best results, with shorter durations limiting optimization and longer durations introducing overfitting risks and additional computational costs.

Loss Weights. In the analysis of loss components, assigning a weight of 0.1 to the reconstruction loss produces the best results. A higher weight overemphasizes feature recovery, which limits the model's ability to learn discriminative representations. A regularization loss weight of 0.9 is most effective for constraining parameter magnitudes, thereby promoting model stability.

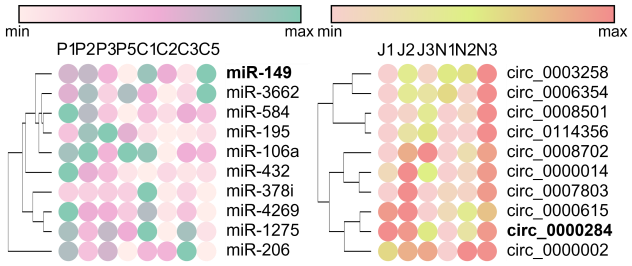

Figure 6: Differential expression analysis of miR-149 and circ_0000284 in HCC.

Case Study

To explore complex disease-related molecular regulatory networks, we conducted a case study on hepatocellular carcinoma (HCC) using the ZeroStem. ZeroStem employs causal convolutions in Mamba to aggregate local relationships in subgraph sequences, enhancing node embeddings and capturing HCC-specific regulatory patterns. We compared the performance of subgraphs defined by four meta-paths (d_c_d, d_m_d, c_d_c, m_d_m) against the entire graph. As shown in Figure 4, removing the m_d_m and d_c_d subgraphs led to substantial performance degradation, underscoring their critical role in CMI prediction.

To further assess their significance, we applied SHAP analysis to quantify the contributions of circRNAs and miRNAs in the m_d_m and d_c_d subgraphs. Representative results, depicted in Figure 5, revealed miR-149 with a contribution value of 0.6824 (ranked 2nd in m_d_m) and circ_0000284 with a contribution value of 0.0602 (ranked 1st in d_c_d). ZeroStem predicted an interaction between them (score: 0.88), suggesting a potential regulatory relationship. Differential expression analysis of GEO datasets, GSE128274 (Sheng et al. 2019) and GSE174179 (Zhao et al. 2025), presented in Figure 6, confirmed significant dysregulation of miR-149 and circ_0000284 in HCC, validating their biological relevance. The findings highlight the effectiveness of ZeroStem's causal convolution architecture for zero-shot CMI prediction, showcasing its capability to elucidate complex regulatory networks and provide a reliable tool for disease mechanism studies.

Conclusion

In this work, we propose ZeroStem, a state space modeling-based dual-channel learning framework for zero-shot CMI prediction. ZeroStem leverages biological prior knowledge to enhance node features and generate meta-path based semantic subgraphs, thereby effectively capturing macro-topological and micro-semantic features. It addresses key challenges in biological networks modeling and zero-shot learning. Experimental results demonstrate that ZeroStem significantly outperforms existing methods across benchmark datasets, offering novel insights into disease biomarker discovery and understanding complex biological systems.

Acknowledgments

This work was supported in part by the Graduate Innovation Program of China University of Mining and Technology, under Grant 2025WLKXJ206, in part by the Fundamental Research Funds for the Central Universities, under Grant 2025-00062, and in part by the Postgraduate Research & Practice Innovation Program of Jiangsu Province, under Grant KYCX25_2893, in part by the Natural Science Foundation of Shandong, under Grant ZR2024MF042, in part by the National Natural Science Foundation of China, under Grants 62573419, 62172355, 61702444, in part by the National Science Foundation for Distinguished Young Scholars of China, under Grant 62325308.

References

Cao, C.; Li, M.; Wang, C.; Xu, L.; Zou, Q.; Wang, Y.; and Han, W. 2025. DGCLCMI: a deep graph collaboration learning method to predict circRNA-miRNA interactions. *BMC biology*, 23(1): 104.

Caporali, A.; Anwar, M.; Devaux, Y.; Katare, R.; Martelli, F.; Srivastava, P. K.; Pedrazzini, T.; and Emanueli, C. 2024. Non-coding RNAs as therapeutic targets and biomarkers in ischaemic heart disease. *Nature reviews cardiology*, 21(8): 556–573.

Clark, K.; Luong, M.-T.; Le, Q. V.; and Manning, C. D. 2020. ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. In *International Conference on Learning Representations*, 3295–3312.

Cui, C.; Zhong, B.; Fan, R.; and Cui, Q. 2024. HMDD v4. 0: a database for experimentally supported human microRNA-disease associations. *Nucleic Acids Research*, 52(D1): D1327–D1332.

Gu, A.; and Dao, T. 2023. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. *arXiv preprint arXiv:2312.00752*.

He, J.; Xu, Q.; Jiang, Y.; Wang, Z.; and Huang, Q. 2024. Ada-gad: Anomaly-denoised autoencoders for graph anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 8481–8489.

Jo, H.; Bu, F.; and Shin, K. 2023. Robust graph clustering via meta weighting for noisy graphs. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 1035–1044.

Lan, W.; Zhu, M.; Chen, Q.; Chen, B.; Liu, J.; Li, M.; and Chen, Y.-P. P. 2020. CircR2Cancer: a manually curated database of associations between circRNAs and cancers. *Database*, 2020: baaa085.

Liu, M.; Wang, Q.; Shen, J.; Yang, B. B.; and Ding, X. 2019. Circbank: a comprehensive database for circRNA with standard nomenclature. *RNA biology*, 16(7): 899–905.

Lu, C.; Han, T.; and Ning, Y. 2022. Context-aware health event prediction via transition functions on dynamic disease graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 4567–4574.

Nejadi Orang, F.; and Abdoli Shadbad, M. 2024. Competing endogenous RNA networks and ferroptosis in cancer: novel therapeutic targets. *Cell Death & Disease*, 15(5): 357.

Nemeth, K.; Bayraktar, R.; Ferracin, M.; and Calin, G. A. 2024. Non-coding RNAs in disease: from mechanisms to therapeutics. *Nature Reviews Genetics*, 25(3): 211–232.

Peng, L.; Wang, W.; Yang, C.; Xiao, W.; Fu, X.; and Chen, Y. 2024. Dual-Stream Heterogeneous Graph Neural Network Based on Zero-Shot Embeddings for Predicting miRNA-Drug Sensitivity. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 1122–1128. IEEE.

Qian, Y.; Zheng, J.; Jiang, Y.; Li, S.; and Deng, L. 2022. Prediction of circRNA-miRNA association using singular value decomposition and graph neural networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 20(6): 3461–3468.

Rozemberczki, B.; Allen, C.; and Sarkar, R. 2021. Multi-scale attributed node embedding. *Journal of Complex Networks*, 9(2): cnab014.

Sheng, Z.; Wang, X.; Xu, G.; Shan, G.; and Chen, L. 2019. Analyses of a panel of transcripts identified from a small sample size and construction of RNA networks in hepatocellular carcinoma. *Frontiers in Genetics*, 10: 431.

Shi, Y.; Huang, Z.; Feng, S.; Zhong, H.; Wang, W.; and Sun, Y. 2021. Masked Label Prediction: Unified Message Passing Model for Semi-Supervised Classification. In Zhou, Z.-H., ed., *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, 1548–1554. International Joint Conferences on Artificial Intelligence Organization. Main Track.

Su, X.; Hu, P.; Li, D.; Zhao, B.; Niu, Z.; Herget, T.; Yu, P. S.; and Hu, L. 2025. Interpretable identification of cancer genes across biological networks via transformer-powered graph representation learning. *Nature Biomedical Engineering*, 1–19.

Sun, Z.-Y.; Yang, C.-L.; Huang, L.-J.; Mo, Z.-C.; Zhang, K.-N.; Fan, W.-H.; Wang, K.-Y.; Wu, F.; Wang, J.-G.; Meng, F.-L.; et al. 2024. circRNADisease v2. 0: an updated resource for high-quality experimentally supported circRNA-disease associations. *Nucleic acids research*, 52(D1): D1193–D1200.

Wang, D.; Cui, P.; and Zhu, W. 2016. Structural deep network embedding. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 1225–1234.

Wang, L.; You, Z.-H.; Li, J.-Q.; and Huang, Y.-A. 2020. IMS-CDA: prediction of CircRNA-disease associations from the integration of multisource similarity information with deep stacked autoencoder model. *IEEE transactions on cybernetics*, 51(11): 5522–5531.

Wang, X.-F.; Huang, L.; Wang, Y.; Guan, R.-C.; You, Z.-H.; Sheng, N.; Xie, X.-P.; and Yang, Q.-X. 2024a. A multichannel graph neural network based on multisimilarity modality hypergraph contrastive learning for predicting unknown types of cancer biomarkers. *Briefings in Bioinformatics*, 25(6): bbae575.

Wang, Z.; Xiong, Z.; Huang, F.; Liu, X.; and Zhang, W. 2024b. ZeroDDI: A Zero-Shot Drug-Drug Interaction Event Prediction Method with Semantic Enhanced Learning and

Dual-modal Uniform Alignment. In Larson, K., ed., *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, 6071–6079. International Joint Conferences on Artificial Intelligence Organization. Main Track.

Wei, M.; Wang, L.; Li, Y.; Li, Z.; Zhao, B.; Su, X.; Wei, Y.; and You, Z. 2024. BioKG-CMI: a multi-source feature fusion model based on biological knowledge graph for predicting circRNA-miRNA interactions. *Science China Information Sciences*, 67(8): 189104.

Wei, M.-M.; Wang, L.; Zhao, B.-W.; Su, X.-R.; You, Z.-H.; and Huang, D.-S. 2025. Integrating transformer and graph attention network for circRNA-miRNA interaction prediction. *IEEE Journal of Biomedical and Health Informatics*, 1–10.

Wu, W.; Zhao, F.; and Zhang, J. 2024. circAtlas 3.0: a gateway to 3 million curated vertebrate circular RNAs based on a standardized nomenclature scheme. *Nucleic Acids Research*, 52(D1): D52–D60.

Yin, W.; Wang, S.; Zhang, Y.; Qiao, S.; Wang, S.; Khan, F.; Altuki, R.; and Lyu, Z. 2025. Autonomously Adjusting Multi-Relational Hypergraphs Structure for Predicting circRNA-MiRNA Associations. *IEEE Journal of Biomedical and Health Informatics*, 1–10.

Yin, W.; Wang, S.; Zhang, Y.; Qiao, S.; Wu, W.; and Li, H. 2024. Multirelational Hypergraph Representation Learning for Predicting circRNA-miRNA Associations. *Journal of Chemical Information and Modeling*, 64(21): 8349–8360.

Zhao, P.; Yin, C.; Liu, R.; Shao, S.; Ke, W.; and Song, Z. 2025. Exosome-Delivered circFOXP1 Upregulates Autophagy and Promotes Hepatocellular Carcinoma Progression Through Its Encoded p196 Protein Targeting the KHDRBS3/ULK1 Axis. *International Journal of Nanomedicine*, 5247–5265.