This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3628673

BiBLDR: Bidirectional Behavior Learning for Drug Repositioning

Renye Zhang, Mengyun Yang, Qichang Zhao, Jianxin Wang

***Abstract*— Many deep learning methods represented by graph-based approaches achieve significant progress in drug repositioning. However, these graph-based methods face a critical limitation: they often fail in cold-start scenarios because the graph structure relies heavily on known association information from both the drug and disease sides. To address this challenge, we propose a bidirectional behavior learning strategy for drug repositioning, BiBLDR, an innovative framework that reformulates drug repositioning as a behavior sequence learning task. First, we construct bidirectional behavioral sequences based on drug and disease sides. Bidirectional behavior sequences ensure sufficient information for model learning in both drug and disease cold-start scenarios, while providing more precise feature representations for association prediction tasks. Subsequently, we propose a two-stage strategy for drug repositioning. In the first stage, we construct prototype spaces to characterise the representational attributes of drugs and diseases. In the second stage, these refined prototypes and bidirectional behavior sequence data are leveraged to predict potential drug-disease associations. This design allows BiBLDR to more robustly capture hidden pharmacological relationships from bidirectional behavioral sequences, delivering significant benefits in cold-start scenarios. Extensive experiments demonstrate that our method achieves state-of-the-art performance on benchmark datasets. Meanwhile, BiBLDR demonstrates significantly superior performance compared to previous methods in cold-start scenarios. Our code is published in https://github.com/Renyeeah/BiBLDR.**

***Index Terms*— Drug repositioning, Drug-disease association prediction, Behavior sequence, Transformer structure**

I. Introduction

DRUG repositioning seeks to identify new therapeutic indications for existing, approved drugs, thereby uncovering potential treatments for diseases that may respond to these agents [1]. The traditional cold-start drug development process is generally divided into three phases, each demanding significant expenditures and prolonged research timelines. Compared to the conventional drug development process, drug repositioning circumvents the need for Phase I clinical trials—which focus on safety evaluation—by utilizing existing clinical safety data from approved drugs. This approach substantially reduces the financial costs and temporal demands of bringing a therapeutic agent to market [2]. The pharmaceutical industry has witnessed numerous clinically and commercially significant successes in drug repositioning, including aspirin [3] (repurposed from an analgesic to a cardiovascular prophylactic), minoxidil [4] (redeveloped from an antihypertensive to treatment for alopecia), and remdesivir [5] (transitioned from an investigational antiviral to therapy for COVID-19).

Computational drug repositioning involves the systematic identification of potential indications for drugs through the development of mathematical models and the design of algorithms, rather than relying solely on incidental discoveries made during clinical practice. BNNR [6] integrates association and similarity matrices and performs matrix completion under the low-rank assumption, which can naturally handle the cold-start problem. LG-DTI [7] learns local representations from drug molecular structures and protein sequences, and uses a semi-supervised isomorph network to obtain global representations. This comprehensive representation method allows each entity to carry richer semantic information. RGLDR [8] uses meta-paths to explore different connection mechanisms in drug repositioning tasks. Each meta-path constructs a regular graph and uses graph representation learning to obtain the embedding of each entity. DDAGDL [9] combines the biological attributes of drugs and diseases with geometric network priors, using geometric deep learning to project non-Euclidean biomedical data into a latent feature space, capturing drug and disease features more comprehensively.

Drug-disease associations and similarity data represent graph topological structures, so many studies conceptualize the drug repositioning task as a link prediction problem within graphs. This contributes to the widespread use of *Graph Convolutional Networks* (GCN) in deep learning-based methods. DRWBNCF [10] uses a weighted bilinear graph convolution operation to integrate association and similarity networks into a unified representation, and uses multilayer perceptrons and graph regularization to model drug-disease associations. PSGCN [11] generates independent subgraphs for each drug-disease pair to be predicted, transforming the link prediction task into a graph classification problem. This approach enhances the model's sensitivity to local topological information. Although the GNN model has achieved significant success, it is prone to over-smoothing simulation when processing

Mengyun Yang is the Corresponding Author. E-mail: mengyun_yang@126.com

Renye Zhang and Mengyun Yang are with the School of Computer Science, Hunan First Normal University, Hunan, China, 410205. Qichang Zhao and Jianxin Wang are with the School of Computer Science and Engineering, Central South University, Hunan, China, 410083.

This work was supported by the National Natural Science Foundation of China (Grant no. 62272309).

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3628673

heterogeneous biological information graphs. To address this issue, iGRLDTI [12] employs a node-dependent local smoothing strategy in heterogeneous bioinformatics graphs to adjust the propagation depth of elements. This strategy significantly enhances the representation capability of entity features and effectively alleviates the problem of over-smoothing.

However, drug repositioning often encounters more challenging scenarios. Some newly approved drugs do not yet have established therapeutic indications, while certain newly identified diseases do not yet have known pharmacological treatments. When they face these cold-start scenarios, existing graph-based methods often yield significantly diminished predictive performance for novel drugs or diseases due to the lack of association information regarding target entities. Advanced graph models achieve an *Area Under the Precision-Recall Curve* (AUPRC) of approximately 0.95 in existing drug or disease repositioning tasks. However, this value drops to around 0.3 when confronted with cold-start scenarios [11], [13]. Section III-E provides a comprehensive analysis of how existing models perform in cold-start scenarios.

To overcome the challenges of cold-starts, consider alternative data structures for modelling the therapeutic relationship between drugs and diseases, rather than relying solely on graph structures. In addition to being abstracted as a graph structure, the associations between drugs and diseases can also be represented as behavioral sequences [14], [15]. Behavioral sequences are extensively utilized in recommendation tasks, referring to the series of actions that users take toward different items [16]–[18]. In drug repositioning tasks, the association information between a drug and multiple diseases can be regarded as the behavior sequence of that drug, while the association information between a disease and multiple drugs can be regarded as the behavior sequence of that disease. In this behavioral sequence–based data structure, the prediction of associations between drugs and diseases no longer relies on a heterogeneous graph of drugs and diseases but instead depends on the behavioral sequences generated by each of them.

The differences between graph-based and behavior sequence-based drug repositioning modelling approaches are illustrated in Fig. 1. Graph-based approaches primarily rely on heterogeneous biological graphs. In cold-start scenarios, the target node becomes isolated due to missing associations, making it difficult for the model to perform inference. In contrast, behavior sequence–based drug repositioning methods primarily rely on bidirectional behavioral sequences between drugs and diseases. Even in cold-start scenarios where the bidirectional sequence degenerates into a unilateral one, the unilateral behavioral sequence still provides sufficient semantic information for the model to learn effectively. Furthermore, in drug repositioning tasks with existing drugs and diseases, modelling the interactions between their bidirectional behavioral sequences provides the model with more precise semantic representations.

Inspired by the aforementioned works, to address the limitations of traditional graph representations in drug repositioning tasks, we propose ***Bi****directional* ***B****ehavior* ***L****earning for* ***D****rug* ***R****epositioning* (BiBLDR), which analyzes drug-disease association data from the perspective of behavioral sequences. In addition, it is the first to apply a Transformer-based architecture to process behavioral sequence data in drug repositioning tasks. Furthermore, our approach is based on a two-stage training process. The first stage uses similarity data to construct prototype spaces for drugs and diseases independently, while the second stage predicts the association between drugs and diseases by analyzing their bidirectional behavioral sequences. The contributions of this paper are as follows:

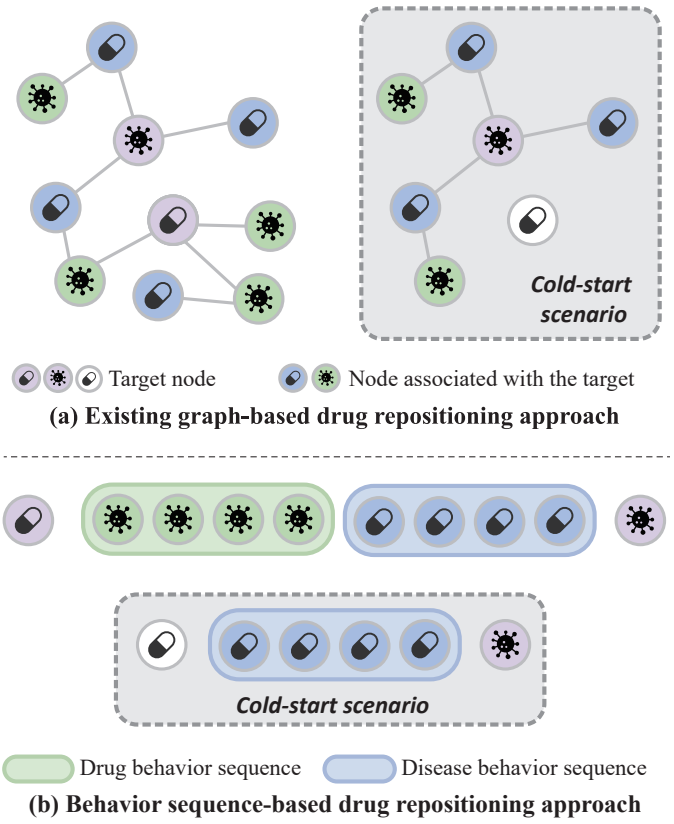


Fig. 1. Comparison of graph-based and behavior sequence-based drug repositioning approaches.

- We propose a novel learning paradigm for drug repositioning: bidirectional behavioral sequence analysis. This strategy simultaneously constructs behavioral sequence data from both drug and disease aspect, leveraging bidirectional information allows the model to effectively mitigate the cold-start problem.
- We propose a two-stage drug repositioning strategy. In the first stage, we employ similarity-driven learning to establish prototype spaces for drugs and diseases. This spatial organization translates entity similarities into geometrically constrained vector relationships, enabling the model to extract structural relationships within the latent prototype space. The subsequent stage integrates these prototypical representations with bidirectional behavioral sequences to predict drug-disease associations, with bidirectional information augmenting feature expressiveness through multi-domain representation fusion.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3628673

- We are the first to apply the Transformer architecture to drug repositioning tasks based on behavioral sequence prediction, utilizing the multi-head attention mechanism to capture dependencies between drug and disease behavioral sequences, which synergistically strengthens the interaction of drug and disease features.
- We conduct extensive experiments on public datasets, and the results show that our BiBLDR significantly outperforms existing methods.

The organization of the remaining sections of this paper is as follows: Section II proposes the construction method of bidirectional behavioral sequences and provides a detailed framework of BiBLDR. Section III describes the datasets, experimental implementation details, and comprehensive results and analysis. Finally, the conclusion is presented in Section IV.

II. Methodology

This section commences with an overview of the methodological framework underpinning BiBLDR, followed by a systematic presentation of the proposed data modelling strategy for bidirectional behavioral sequences in computational drug repositioning. Subsequently, a comprehensive exposition of the framework is presented, detailing its architectural components, operational mechanisms, and theoretical underpinnings.

A. Overall Framework

Unlike conventional graph-based representation learning approaches, our methodology redefines drug repositioning as a behavioral sequence recommendation task. Our framework transmutes static matrix structures into behavioral sequences and proposes a two-stage learning strategy to perform drug repositioning task. In the first stage, drugs and diseases are represented as feature vectors derived from their similarity sequences. The Siamese neural network refines these representations by minimizing discrepancies between predicted and actual similarity scores, generating optimized prototype spaces for both entities. The second stage processes behavioral sequences of drugs and diseases in parallel, integrating similarity information to learn enriched semantic representations from bidirectional behavior sequences. Simultaneously, domain-specific attribute features are extracted through specialized pipelines. The entity-attribute and bidirectional behavioral representations are integrated to generate the final association prediction scores. Fig. 2 provides an overview of the overall framework.

B. Bidirectional Behavioral Sequence modelling

Traditional behavioral sequence prediction tasks focus on unidirectional user-item interactions. By contrast, our work jointly models the behavioral sequences of drugs and diseases to capture their bidirectional relationships. Specifically, our framework accounts for a drug's therapeutic effects on diseases and a disease's sensitivity to drugs. This bidirectional modelling enhances sequence representation by integrating drug-disease interdependencies. This section provides a detailed description of the construction method for the drug-disease bidirectional behavior sequence.

1) Dataset Format: The dataset used contains a drug-disease association matrix $\mathcal{A} \in \mathbb{R}^{|\mathcal{U}|\times|\mathcal{V}|}$, a drug-drug similarity matrix $\mathcal{S}^{\mathcal{U}} \in \mathbb{R}^{|\mathcal{U}|\times|\mathcal{U}|}$, a disease-disease similarity matrix $\mathcal{S}^{\mathcal{V}} \in \mathbb{R}^{|\mathcal{V}|\times|\mathcal{V}|}$. $|\mathcal{U}|$ and $|\mathcal{V}|$ represent the total number of drugs and diseases. $\mathcal{A}_{ij} \in \{0,1\}$, indicating the therapeutic relationship between i-th drug and j-th disease. All elements of $\mathcal{A}$ are divided into the training set and test set, with only the elements in the training set being accessible. $\mathcal{S}^{\mathcal{U}}_{ij} \in [0,1]$, indicating the similarity between i-th and j-th drug. $\mathcal{S}^{\mathcal{V}}_{ij} \in [0,1]$, indicating the similarity between i-th and j-th disease.

2) Problem Setting: We define a scenario that uses behavioral sequences to model the drug repositioning. Let $\mathcal{U} = \{u_1, u_2, \cdots, u_{|\mathcal{U}|}\}$ and $\mathcal{V} = \{v_1, v_2, \cdots, v_{|\mathcal{V}|}\}$. $\mathcal{U}$ and $\mathcal{V}$ represent the sets of all drugs and diseases in the dataset, respectively. u_i and v_j represent the i-th drug and j-th disease. Predicting the therapeutic effect of u_k on disease v_m is essentially predicting the value of $\mathcal{A}_{km}$. Since $\mathcal{A}_{km} \in \{0,1\}$, the drug reposition task can be treated as a binary classification task. $\mathcal{A} = \{\mathcal{A}^{train}, \mathcal{A}^{test}\}$ represent the training set and test set, respectively. The complete content of the proposed bidirectional behavioral sequence is $(u_k, v_m, \mathcal{R}_k, \mathcal{I}_m, \mathcal{A}_{km})$, includes *drug*, *disease*, *drug behavioral sequence*, *disease behavioral sequence*, and *label*. For the $\mathcal{A}_{km}$ prediction task, the drug and disease are denoted as u_k and v_m, respectively. The drug behavior sequence is composed of diseases related to u_k in the training set, excluding v_m. It can be defined as $\mathcal{R}_k = \{v_j \mid 1 \le j \le |\mathcal{V}|, j \ne m, \mathcal{A}_{kj} \in \mathcal{A}^{train}\}$. The disease behavior sequence is composed of drugs related to v_m in the training set, excluding u_k. It can be defined as $\mathcal{I}_m = \{u_i \mid 1 \le i \le |\mathcal{U}|, i \ne k, \mathcal{A}_{im} \in \mathcal{A}^{train}\}$. For the prediction task of u_k and v_m, the label is $\mathcal{A}_{km}$.

C. Constructing Latent Prototype Representations

In behavior sequence prediction tasks, the intrinsic attributes of users and items serve as vital characterizing information, enabling a more comprehensive understanding of entities. To fully leverage the intrinsic attributes of drugs and diseases, we construct latent prototype representations for drugs and diseases using structural relationships derived from similarity data. These prototypes are structured into separate drug and disease prototype spaces to retain domain-specific semantics. As shown in Fig. 2(a), initial representations from similarity matrices are refined via a Siamese-based network, and cosine similarity aligns the prototype spaces.

1) Generating Initial Representations: The initial representation for each drug or disease is derived from its corresponding similarity matrix by isolating its full relational profile—formally equivalent to extracting either a row or column vector from the drug-drug ($\mathcal{S}^{\mathcal{U}}$) or disease-disease ($\mathcal{S}^{\mathcal{V}}$) similarity matrix, respectively. This vector extraction process preserves the complete similarity structure of each entity within its domain and provides geometrically interpretable input for subsequent prototype refinement.

2) Siamese Structures for Feature Learning: The initial similarity-based representations are processed through a Siamese network architecture [19] to refine their respective prototype vectors iteratively. The Siamese network comprises

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3628673

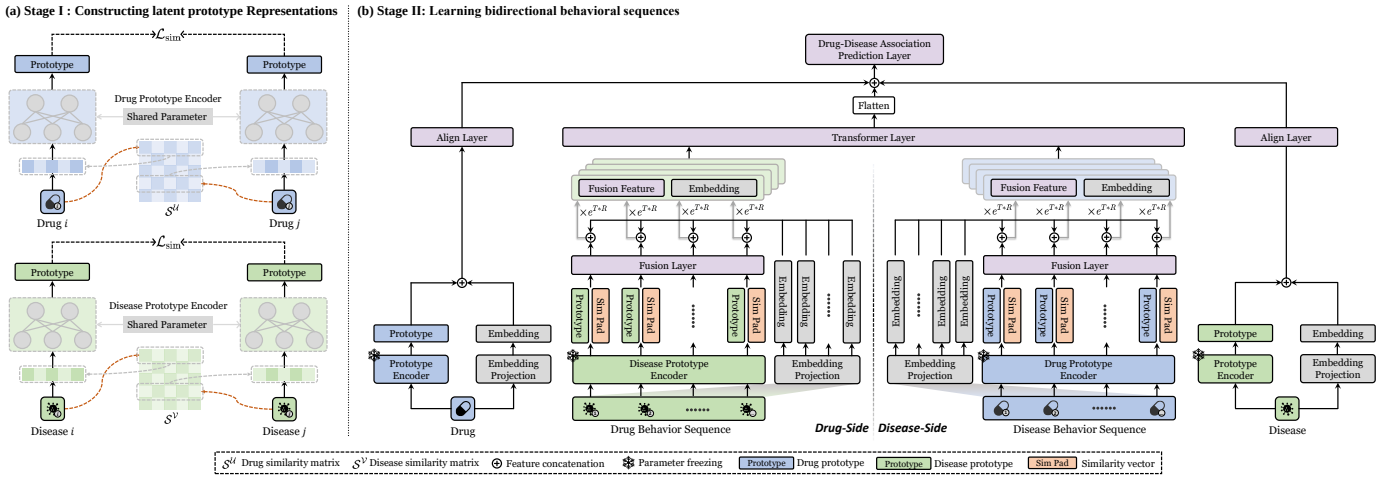


Fig. 2. The proposed BiBLDR framework. (a) Utilize similarity data to construct prototype spaces for drugs and diseases separately. (b) Utilize prototypes and bidirectional behavioral sequence information to predict drug-disease associations.

two parallel feature encoders with identical topological structures and shared weight parameters, ensuring parameter-space symmetry during joint optimization. This enhances the consistency of intra-domain feature transformation and reduces computational overhead compared to using separate encoders for feature extraction.

3) Prototype Alignment on Latent Space: We use cosine similarity to align the prototype space generated by the encoder. The calculation method for the cosine similarity between prototypes is shown in Eq. (1):

$$\mathrm{sim}(\mathbf{P}_{u_i}, \mathbf{P}_{u_j}) = \frac{\mathbf{P}_{u_i} \cdot \mathbf{P}_{u_j}}{\|\mathbf{P}_{u_i}\| \cdot \|\mathbf{P}_{u_j}\|} \tag{1}$$

where $\mathbf{P}u_i$ and $\mathbf{P}u_j$ represent the outputs of the Siamese network for entities u_i and u_j, respectively, and $|.|$ represents the L^2-norm, cosine similarity encodes angular relationships that map drug-disease affinities onto a geometrically interpretable manifold. By focusing on direction rather than magnitude, it separates semantic relatedness from spurious variations, embedding similarities as angular proximity in latent space. To construct latent prototype spaces, the Siamese architecture is optimized via the contrastive loss defined in Eq. (2).

$$\begin{aligned} \mathcal{L}_{\mathrm{sim}}^{\mathcal{U}} &= \sum_{i=1}^{|\mathcal{U}|} \sum_{j=i+1}^{|\mathcal{U}|} \left(\mathcal{S}_{ij}^{\mathcal{U}} - \mathrm{sim}\left(f_\phi^{\mathcal{U}}(u_i), f_\phi^{\mathcal{U}}(u_j)\right)\right)^2 \\ \mathcal{L}_{\mathrm{sim}}^{\mathcal{V}} &= \sum_{i=1}^{|\mathcal{V}|} \sum_{j=i+1}^{|\mathcal{V}|} \left(\mathcal{S}_{ij}^{\mathcal{V}} - \mathrm{sim}\left(f_\phi^{\mathcal{V}}(v_i), f_\phi^{\mathcal{V}}(v_j)\right)\right)^2 \end{aligned} \tag{2}$$

where $f_\phi^{\mathcal{U}}$ and $f_\phi^{\mathcal{V}}$ represent the Siamese networks processing drug data and disease data, respectively. while $f_\phi^{\mathcal{U}}(u_i)$ and $f_\phi^{\mathcal{V}}(v_i)$ are equivalent to the extracted prototypes $\mathbf{P}_{u_i}$ and $\mathbf{P}_{v_i}$. This dual-stream refinement mechanism ensures that drug and disease prototypes capture domain-specific semantics with high fidelity while maintaining computational efficiency.

D. Learning Bidirectional Behavioral Sequences

In this stage, BiBLDR utilizes the drug and disease prototypes generated in the first phase to further process bidirectional behavioral sequences. Fig. 2(b) illustrates the main pipeline of the model. The methodological workflow can be systematically decomposed into four discrete computational phases: embedding projection, similarity information fusion, attention mechanism for sequence processing, and drug-disease association prediction.

1) Embedding Projection: In addition to generating prototypes for drugs and diseases using similarity information in stage I, to further enrich the semantic information inherently embedded within drugs and diseases, through a structured embedding framework, each ID is mapped to a low-dimensional vector representation, termed an embedding, which is iteratively refined through optimization during model training. More precisely, two dedicated learnable matrices, $\mathcal{W}^{\mathcal{U}} \in \mathbb{R}^{|\mathcal{U}| \times d_w}$ and $\mathcal{W}^{\mathcal{V}} \in \mathbb{R}^{|\mathcal{V}| \times d_w}$, are initialized to encode trainable embeddings for all $|\mathcal{U}|$ drugs and $|\mathcal{V}|$ diseases, respectively, where d_w denotes the dimensionality. These embeddings are incorporated into subsequent computational layers, enabling the model to capture relational properties through continual optimization, thereby distilling parametric relationships and latent interdependencies inherent in behavior sequence data.

2) Similarity Information Fusion: Although the entity prototypes derived from stage I training implicitly encapsulate similarity information, this remains insufficient. In this stage, we explicitly integrate the numerical similarity values between entities directly into the prototypes within the behavioral sequences. This integration amplifies the distinction in features among entities with varying degrees of similarity, rendering their differences more pronounced. For the prediction task $\mathcal{A}_{km}$, the constructed bidirectional behavioral sequence consists of the drug-side behavior sequence $\mathcal{R}_k$ and the disease-side behavioral sequence $\mathcal{I}_m$. Initially, each drug and disease in $\mathcal{I}_m$ and $\mathcal{R}_k$, respectively, is embedded into the latent prototype space using the pre-trained prototype encoders from

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3628673

the Stage I, as shown in Eq. (3).

$$\begin{aligned}\mathcal{P}_k^{\mathcal{U}} &= \{\mathbf{P}_{v_i} \mid \mathbf{P}_{v_i} = f_\phi^{\mathcal{V}}(v_i), v_i \in \mathcal{R}_k^b\} \\ \mathcal{P}_m^{\mathcal{V}} &= \{\mathbf{P}_{u_i} \mid \mathbf{P}_{u_i} = f_\phi^{\mathcal{U}}(u_i), u_i \in \mathcal{I}_m^b\}\end{aligned} \tag{3}$$

where $\mathbf{P}_{u_i}$ and $\mathbf{P}_{v_i}$ represent the extracted prototypes from u_i and v_i. Similarity information is obtained from elements in the behavior sequence and target elements. For u_i, the similarity matrix $\mathcal{S}^{\mathcal{U}}$ is used to obtain its similarity $\mathcal{S}_{ki}^{\mathcal{U}}$ with the target drug u_k. For v_i, its similarity $\mathcal{S}_{mi}^{\mathcal{V}}$ with the target disease v_m is obtained. To integrate entity similarity into prototypes, we convert numerical similarity scores between each entity in the behavioral sequence and the target entity into vectors matching the prototype's dimensions. The padding method applied to these vectors is detailed in Eq. (4).

$$\begin{aligned}\mathbf{S}_{ki}^{\mathcal{U}} &= \mathcal{S}_{ki}^{\mathcal{U}} \cdot \mathbf{1} \\ \mathbf{S}_{mi}^{\mathcal{V}} &= \mathcal{S}_{mi}^{\mathcal{V}} \cdot \mathbf{1}\end{aligned} \tag{4}$$

where $\mathbf{S}_{ki}^{\mathcal{U}}$ represents the similarity vector generated between drug u_k and u_i, $\mathbf{S}_{mi}^{\mathcal{V}}$ represents the similarity vector generated between disease v_m and v_i, and $\mathbf{1}$ is a vector of ones with a dimension of d_0. To synthesize prototype vectors and similarity vectors for entities, the fusion layer combines the prototype with its corresponding similarity vector and then outputs a fusion vector with dimension d_0. This integration ensures that both intrinsic behavioral patterns and explicit similarity relationships are preserved in the feature representation. The merged feature vector is computed as shown in Eq. (5):

$$\begin{aligned}\mathcal{H}_{km}^{\mathcal{U}} &= \{\mathbf{H}_{mi}^{\mathcal{U}} \mid \mathbf{H}_{mi}^{\mathcal{U}} = f_\varphi^{\mathcal{U}}(\mathbf{P}_{v_i} \oplus \mathbf{S}_{mi}^{\mathcal{V}}), \mathbf{P}_{v_i} \in \mathcal{P}_k^{\mathcal{U}}\} \\ \mathcal{H}_{km}^{\mathcal{V}} &= \{\mathbf{H}_{ki}^{\mathcal{V}} \mid \mathbf{H}_{ki}^{\mathcal{V}} = f_\varphi^{\mathcal{V}}(\mathbf{P}_{u_i} \oplus \mathbf{S}_{ki}^{\mathcal{U}}), \mathbf{P}_{u_i} \in \mathcal{P}_m^{\mathcal{V}}\}\end{aligned} \tag{5}$$

where $\oplus$ represents the feature concatenation operation, $f_\varphi^{\mathcal{U}}$ and $f_\varphi^{\mathcal{V}}$ are the fusion layers on drug-side and disease-side, respectively. In our model, the fusion layer comprises a feedforward neural network. $\mathbf{H}_{ki}^{\mathcal{U}}$ and $\mathbf{H}_{mj}^{\mathcal{V}}$ are the fused features, with a dimensional size of d_0.

3) Attention Mechanism for Sequence Process: Following the fusion of similarity information into the prototypes of entities within bidirectional behavioral sequences, the entity embeddings from $\mathcal{W}^{\mathcal{U}}$ and $\mathcal{W}^{\mathcal{V}}$ are concatenated with similarity-fused feature vectors to enhance representational capacity, while ratings in the behavior sequence is also incorporated. Unlike conventional methods that directly multiply feature vectors and ratings, our approach applies a logarithmic transformation to binarized ratings before interacting with feature vectors. This preserves semantic integrity of negative samples for contrastive learning. Eq. (6) formalizes this operation.

$$\begin{aligned}\mathcal{N}_{km}^{\mathcal{U}} &= \{\mathbf{X}_{mi}^{\mathcal{U}} \mid \mathbf{X}_{mi}^{\mathcal{U}} = \left(\mathbf{H}_{mi}^{\mathcal{U}} \oplus \mathcal{W}_i^{\mathcal{V}}\right) e^{\mathcal{T} \cdot \mathcal{A}_{ki}}, \mathbf{H}_{mi}^{\mathcal{U}} \in \mathcal{H}_{km}^{\mathcal{U}}\} \\ \mathcal{N}_{km}^{\mathcal{V}} &= \{\mathbf{X}_{ki}^{\mathcal{V}} \mid \mathbf{X}_{ki}^{\mathcal{V}} = \left(\mathbf{H}_{ki}^{\mathcal{V}} \oplus \mathcal{W}_i^{\mathcal{U}}\right) e^{\mathcal{T} \cdot \mathcal{A}_{im}}, \mathbf{H}_{ki}^{\mathcal{V}} \in \mathcal{H}_{km}^{\mathcal{V}}\}\end{aligned} \tag{6}$$

where $\mathcal{T}$ is a tunable hyperparameter, and $\mathcal{W}_i^{\mathcal{U}}$ and $\mathcal{W}_i^{\mathcal{V}}$ represent the i-th row of $\mathcal{W}^{\mathcal{U}}$ and the i-th row of $\mathcal{W}^{\mathcal{V}}$, which correspond to the embedding of the i-th disease and the embedding of the i-th drug, respectively. $\mathcal{N}_{km}^{\mathcal{U}}$ and $\mathcal{N}_{km}^{\mathcal{V}}$ represent the sequence sets on the drug-side and disease-side, respectively, which are the inputs to the Transformer layer. Following [20], we use a Transformer layer based on a multi-head self-attention mechanism to process the bidirectional behavioral sequence features $\mathcal{N}_{km}^{\mathcal{U}}$ and $\mathcal{N}_{km}^{\mathcal{V}}$. For ease of understanding, $\mathcal{N}_{km} = \{\mathcal{N}_{km}^{\mathcal{U}}, \mathcal{N}_{km}^{\mathcal{V}}\}$ is simplified as $\mathbf{X} \in \mathbb{R}^{L_{\text{seq}} \times d_1}$ when describing the forward propagation process of the Transformer layer. $d_1 = d_w + d_0$, L_{seq} represents the aggregate count of all sequences contained within $\mathcal{N}_{km}^{\mathcal{U}}$ and $\mathcal{N}_{km}^{\mathcal{V}}$. The structural diagram of the Transformer layer is shown in Fig. 3.

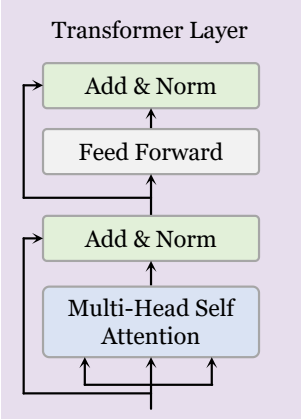


Fig. 3. The structure of the Transformer layer includes the multi-head self-attention mechanism, normalization layer, feedforward neural network, and residual connections.

A Transformer layer has a total of n attention heads, the input matrix $\mathbf{X}$ will be projected into n different low-dimensional subspaces by multiplying learnable weight matrices and then scaled dot-product attention will be computed, as shown in Eq. (7).

$$\begin{aligned}\text{head}_i &= \text{Attention}(\mathbf{X}\mathbf{W}_i^Q, \mathbf{X}\mathbf{W}_i^K, \mathbf{X}\mathbf{W}_i^V) \\ \text{Attention}\left(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i\right) &= \text{softmax}\left(\frac{\mathbf{Q}_i \mathbf{K}_i^T}{\sqrt{d_k}}\right)\mathbf{V}_i\end{aligned} \tag{7}$$

where $\mathbf{Q}_i$ represents the queries of the i-head, $\mathbf{K}_i$ represents the keys, and $\mathbf{V}_i$ represents the values, these values are derived by computing the product of $\mathbf{X}$ and each respective learned weight matrix ($\mathbf{W}_i^Q, \mathbf{W}_i^K, \mathbf{W}_i^V \in \mathbb{R}^{d_1 \times d_k}$). $d_k = d_1/n$, $\sqrt{d_k}$ is used to scale the scores to mitigate gradient vanishing. The outputs from all attention heads are concatenated and subsequently multiplied by a trainable weight matrix, as shown in Eq. (8).

$$\text{MH}(\mathbf{X}) = \text{Concat}(\text{head}_1, \cdots, \text{head}_n) \cdot \text{W}^O \tag{8}$$

where $\mathbf{W}^O$ is a trainable weight matrix, Concat denotes the feature concatenation operation, and $\text{MH}(\mathbf{X})$ represents the output of the multi-head self-attention mechanism. The outputs of subsequent layers are shown in Eq (9).

$$\begin{aligned}\mathbf{O} &= \text{BN}(\mathbf{X} + \text{MH}(\mathbf{X})) \\ \mathbf{O}' &= \text{BN}(\mathbf{O} + \text{FFN}(\mathbf{O}))\end{aligned} \tag{9}$$

where BN denotes the batch normalization layer, FFN represents a feed-forward neural network responsible for endowing the Transformer layer with nonlinear transformation capability. $\mathbf{O}'$ is the final output of the Transformer layer, which can be expressed as $\mathcal{O}_{km}^{\text{TL}} \in \mathbb{R}^{L_{\text{seq}} \times d_1}$.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3628673

4) Drug-disease Association Prediction: In addition to the bilateral behavioral sequence features, we integrate the intrinsic attributes of the drug u_k and the disease v_m into the final prediction of the association A_{km}. Specifically, we extract the prototypes and embeddings of u_k and v_m, which are then concatenated with the bilateral behavioral sequence features $\mathcal{O}_{km}^{\mathrm{TL}}$ to form the input for the association prediction model. To ensure semantic consistency, the prototypes and embeddings are first processed through an alignment layer, which harmonizes their semantic information with the output of the Transformer layer before concatenation with $\mathcal{O}_{km}^{\mathrm{TL}}$. This progress is shown in Eq. (10).

$$\mathbf{M}_{km} = f_{\alpha}^{\mathcal{U}}(\mathbf{P}_{u_k} \oplus \mathcal{W}_k^{\mathcal{U}}) \oplus \mathcal{O}_{km}^{\mathrm{TL}} \oplus f_{\alpha}^{\mathcal{V}}(\mathbf{P}_{v_m} \oplus \mathcal{W}_m^{\mathcal{V}}) \quad (10)$$

where $f_{\alpha}^{\mathcal{U}}$ and $f_{\alpha}^{\mathcal{V}}$ represent the alignment layers for the drug and disease sides, respectively, each constructed using feed-forward neural networks. The association prediction model processes a condensed one-dimensional feature vector, representing comprehensively encoded drug-disease interaction patterns, to generate a scalar output quantifying the predicted therapeutic relationship. This component is architecturally implemented as a multilayer feedforward neural network. In Stage II optimization, a task-specific binary cross-entropy loss function is used to jointly optimize all model components, refining parameters via backpropagation for full convergence. The loss function is shown in Eq. (11).

$$\begin{aligned}\mathcal{L}_{BCEWL} = &\frac{1}{N}\sum_{i=1}^{N}(y_i \cdot \log(\sigma(p_i)) \\ &+ (1-y_i)\cdot\log(1-\sigma(p_i)))\end{aligned} \quad (11)$$

where N represents the total number of training samples, σ denotes the Sigmoid function, each sample x_i has a binary label $y_i \in \{0,1\}$, and p_i represents the predicted probability of sample x_i being classified as positive. In summary, the comprehensive workflow of our proposed bidirectional behavioral sequence learning strategy is formally outlined in Algorithm 1.

Algorithm 1: Bidirectional behavior learning on BiBLDR

Input: test dataset $\mathcal{A}^{test}$, promote matrix $\mathcal{W}^{\mathcal{U}}$ and $\mathcal{W}^{\mathcal{V}}$, prototype encoder $f_{\phi}^{\mathcal{U}}$ and $f_{\phi}^{\mathcal{V}}$, trained model (fusion layer $f_{\varphi}^{\mathcal{U}}$ and $f_{\varphi}^{\mathcal{V}}$, Transformer layer τ, align layer $f_{\alpha}^{\mathcal{U}}$ and $f_{\alpha}^{\mathcal{V}}$, and association prediction layer f_{δ}).
Output: Predicted $\mathcal{A}^{test}$
for $\mathcal{A}_{km} \in \mathcal{A}^{test}$ **do**
  /* Drug-side sequence process */
  $\mathcal{R}_k \leftarrow$ Get drug-side behavior sequence of u_k;
  for $v_i \in \mathcal{R}_k$ **do**
    $\mathbf{P}_{v_i} \leftarrow f_{\phi}^{\mathcal{V}}(v_i)$;
    $\mathbf{S}_{mi}^{\mathcal{V}} \leftarrow \mathcal{S}_{mi}^{\mathcal{V}} \cdot \mathbf{1}$;
    $\mathbf{H}_{mi}^{\mathcal{U}} \leftarrow f_{\varphi}^{\mathcal{U}}(\mathbf{P}_{v_i} \oplus \mathbf{S}_{mi}^{\mathcal{V}})$;
    $\mathbf{X}_{mi}^{\mathcal{U}} \leftarrow \left(\mathbf{H}_{mi}^{\mathcal{U}} \oplus \mathcal{W}_i^{\mathcal{V}}\right) e^{\mathcal{T}\cdot\mathcal{A}_{ki}}$;
    $\mathcal{N}_{km}^{\mathcal{U}}$ append $\mathbf{X}_{mi}^{\mathcal{U}}$;
  end
  /* Disease-side sequence process */
  $\mathcal{I}_m \leftarrow$ Get disease-side behavior sequence of v_m;
  for $u_i \in \mathcal{I}_m$ **do**
    $\mathbf{P}_{u_i} \leftarrow f_{\phi}^{\mathcal{U}}(u_i)$;
    $\mathbf{S}_{ki}^{\mathcal{U}} \leftarrow \mathcal{S}_{ki}^{\mathcal{U}} \cdot \mathbf{1}$;
    $\mathbf{H}_{ki}^{\mathcal{V}} \leftarrow f_{\varphi}^{\mathcal{V}}(\mathbf{P}_{u_i} \oplus \mathbf{S}_{ki}^{\mathcal{U}})$;
    $\mathbf{X}_{ki}^{\mathcal{U}} \leftarrow \left(\mathbf{H}_{ki}^{\mathcal{V}} \oplus \mathcal{W}_i^{\mathcal{U}}\right) e^{\mathcal{T}\cdot\mathcal{A}_{im}}$;
    $\mathcal{N}_{km}^{\mathcal{V}}$ append $\mathbf{X}_{ki}^{\mathcal{V}}$;
  end
  $\mathcal{N}_{km} \leftarrow \{\mathcal{N}_{km}^{\mathcal{U}}, \mathcal{N}_{km}^{\mathcal{V}}\}$;
  $\mathcal{O}_{km}^{\mathrm{TL}} \leftarrow \tau(\mathcal{N}_{km})$;
  $\mathbf{M}_{km} \leftarrow f_{\alpha}^{\mathcal{U}}(\mathbf{P}_{u_k} \oplus \mathcal{W}_k^{\mathcal{U}}) \oplus \mathcal{O}_{km}^{\mathrm{TL}} \oplus f_{\alpha}^{\mathcal{V}}(\mathbf{P}_{v_m} \oplus \mathcal{W}_m^{\mathcal{V}})$;
  $\mathcal{A}_{km} \leftarrow f_{\delta}(\mathbf{M}_{km})$
end

III. Experiments And Discussion

In this section, we first describe the dataset used, the evaluation metrics, and the experimental implementation details. Next, we perform 10-fold cross-validation experiments using BiBLDR with other baseline models on the benchmark datasets to validate the performance achieved by BiBLDR. Ablation experiments are also configured to verify the contribution of individual components in BiBLDR. Additionally, we conduct cold-start experiments, sparse environment testing, and further analysis. These experiments validate BiBLDR's capability in identifying new drug indications, adaptability to sparse data, and the impact of hyperparameter settings on performance. Finally, we conduct case studies to validate the practical applicability of BiBLDR.

A. Datasets and Evaluation Metrics

To comprehensively evaluate BiBLDR's performance, we tested it on three benchmark datasets: Gdataset [21], Cdataset [22], and LRSSL [23]. The Gdataset comprises 593 drugs sourced from DrugBank [24] and 313 diseases from *Online Mendelian Inheritance in Man* (OMIM) [25], with 1,933 experimentally validated drug-disease associations. The Cdataset includes 633 drugs and 409 diseases, featuring 2,352 validated drug-disease associations curated from the Comparative Toxicogenomics Database [26]. Lastly, the LRSSL dataset contains 763 drugs and 681 diseases, encompassing 3,051 known drug-disease interactions.Following previous research [11], [27], drug similarity is calculated using 2D chemical fingerprints. The molecular structures of drugs are obtained from DrugBank, and fingerprints are calculated using the default parameters of the Chemical Development Kit [28]. Similarity is then calculated using the Tanimoto score [29]. The similarity of diseases is obtained by calculating the similarity of disease phenotypes using MimMiner [30], which measures it by calculating the similarity between MeSH terms [31] in OMIM descriptions. Table I provides an overview of the key characteristics and statistical details for these datasets.We select AUPRC and *Area Under the Receiver Operating Characteristic Curve* (AUROC) as the primary evaluation metrics to evaluate model performance. These metrics are widely adopted in drug repositioning tasks due to their robustness and relevance. In addition, we separately compute

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3628673

Recall, Accuracy, and *Normalized Discounted Cumulative Gain* (NDCG) to evaluate model performance from multiple perspectives, thereby providing a more comprehensive assessment of BiBLDR's capabilities. Recall focuses on the number of positive samples correctly identified by the model, Accuracy focuses on the model's prediction results for all test samples, and NDCG focuses on whether the model assigns sufficiently high prediction scores to distinguish positive samples from negative samples.

TABLE I

STATISTIC INFORMATION ABOUT BENCHMARK DATASETS. **NO. OF DRUGS**, **NO. OF DISEASES**, AND **NO. OF ASSOCIATIONS** REPRESENT THE TOTAL NUMBER OF DRUGS, DISEASES, AND KNOWN DRUG-DISEASE ASSOCIATIONS, RESPECTIVELY. **SPARSITY** REPRESENTS THE DEGREE OF SPARSITY IN THE DRUG-DISEASE ASSOCIATION MATRIX.

Dataset	No. of drugs	No. of diseases	No. of associations	Sparsity
Gdataset	593	313	1933	0.0104
Cdataset	663	409	2532	0.0093
LRRSL	269	598	18416	0.1145

B. Implementation Details

To ensure the rigor of performance evaluation, we report model performance using 10-fold cross-validation. We first divide all positive samples in the drug-disease association matrix $\mathcal{A}$ into 10 equal folds $\{\mathcal{A}_1^{\text{pos}}, \mathcal{A}_2^{\text{pos}}, \cdots, \mathcal{A}_{10}^{\text{pos}}\}$. An equal number of negative samples are randomly selected from $\mathcal{A}$ and split into 10 corresponding folds $\{\mathcal{A}_1^{\text{neg}}, \mathcal{A}_2^{\text{neg}}, \cdots, \mathcal{A}_{10}^{\text{neg}}\}$. These folds are paired to create 10 cross-validation subsets $\mathcal{T}_{\text{cv}} = \{(\mathcal{A}_i^{\text{pos}}, \mathcal{A}_i^{\text{neg}}) \mid 1 \leq i \leq 10\}$. During evaluation, we rotate one fold as the test set and use the remaining nine for training, repeating this until all folds have served as the test set. To ensure statistical reliability, the experiment is repeated 10 times independently, with final results averaged across all runs. We employ AdamW as the optimizer for the neural network and utilize a cosine annealing scheduler for learning rate adjustment. During stage I, the learning rate is set to 0.01, while in stage II, it is reduced to 0.0001. The implementation is developed using Python 3.8.19 and PyTorch 2.4.0, and all experiments are conducted on an NVIDIA GeForce RTX 4070 Laptop GPU.

C. Comparative Analysis with Baselines

To validate BiBLDR's state-of-the-art performance, we benchmark it against diverse baseline models on benchmark datasets. The baselines encompass non-deep learning methods: SCMFDD [32], iDrug [33], BNNR [6], and deep learning approaches: NIMCGAN [34], PSGCN [11], DVGEDR [35], SAGAN [13], and MDGCN [36]. This selection ensures robust validation across traditional and modern drug repositioning methodologies. Results from Table II demonstrate that BiBLDR achieves the best performance across all evaluated metrics. On the Gdataset, Cdataset, and LRSSL, BiBLDR attains AUROC of 0.9941, 0.9978, and 0.9950, surpassing the second-best model, MDGCN, by margins of 0.0368, 0.0311, and 0.0320, respectively. Similarly, its AUPRC reach 0.9950, 0.9982, and 0.9964 on these datasets, exceeding the MDGCN by 0.0321, 0.0263, and 0.0271. These results underscore BiBLDR's exceptional capability for drug repurposing. Furthermore, BiBLDR's Recall of 0.9753, 0.9826, and 0.9717 on the three datasets outperform the second-best model by over 0.06. This substantial improvement highlights its superior ability to identify positive samples compared to all baseline models. The overall accuracy scores of 0.9688, 0.9821, and 0.9711 also remain significantly higher than those of other models. For NDCG, BiBLDR consistently achieves values above 0.999 across all datasets, indicating that the results recommended by BiBLDR based on behavioral sequences remain largely consistent with the information in the original data. In conclusion, the experimental results confirm that BiBLDR's performance is significantly superior to all other baseline methods, including advanced GCNs. MDGCN was published in 2025. Building upon GCN, it introduces cross-view and cross-layer contrastive learning to effectively address the weakly supervised problem in drug-disease connections by aligning node embeddings across different views.

D. Ablation Studies

To assess the contribution of individual modules in BiBLDR, we perform ablation studies on its components. The detailed results are summarized in Table III. [A] signifies the exclusion of the entity's prototype, relying solely on the embedding vector to represent the entity's attributes. The results show that the prototype space for drugs and diseases enhances the intrinsic attributes of entities. [B] and [C] represent the use of only disease-side and drug-side behavioral sequence data, respectively, for model training. The results show that the absence of behavioral sequences on either side severely impairs the model's performance. When bidirectional behavioral sequences degenerate into unidirectional behavioral sequences, it affects the model's ability to capture the association between drugs and diseases. This indicates that bidirectional behavioral sequences can more accurately model the therapeutic relationship between drugs and diseases. Particularly, when disease-side data is missing, AUROC and AUPRC experience a sharp decline. On the one hand, this is attributed to the fact that the disease-side behavioral sequence, composed of drugs, can provide more training samples than the drug-side behavioral sequence, composed of diseases, because there are more drugs than diseases in the data set used. Furthermore, this phenomenon also shows that, compared with similarities between diseases, similarities between drugs carry richer semantic information and represent more useful data patterns. [D] demonstrates that the similarity information fusion strategy in stage II effectively enhances the representational capability of prototypes. [E] illustrates that applying the multi-head self-attention mechanism enables BiBLDL to capture the interactive relationships among multiple behavioral sequence features more accurately.

As shown in Table III, from a model design perspective, the similarity fusion mechanism exerts a significant impact on overall performance. Therefore, in addition to general component ablation experiments, we compare different similarity

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3628673

TABLE II
A COMPARATIVE EXPERIMENT INVOLVING 10 TIMES 10-FOLD CROSS-VALIDATION WITH OTHER MODELS ON GDATASET, CDATASET, AND LRSSL. THE HIGHEST-PERFORMING METRIC IS DENOTED IN BOLD.

Metric	Dataset	SCMFDD	iDrug	BNNR	NIMCGCN	PSGCN	DVGEDR	SAGAN	MDGCN	BiBLDR (Ours)
AUROC	Gdataset	0.7731	0.9078	0.9412	0.8234	0.9485	0.9454	0.9518	0.9573	**0.9941**
	Cdataset	0.7896	0.9294	0.9522	0.8393	0.9566	0.9635	0.9643	0.9667	**0.9978**
	LRSSL	0.7698	0.8993	0.9201	0.7581	0.9395	0.9468	0.9500	0.9630	**0.9950**
AUPRC	Gdataset	0.7749	0.9265	0.9575	0.8590	0.9558	0.9525	0.9575	0.9629	**0.9950**
	Cdataset	0.7878	0.9454	0.9654	0.8728	0.9627	0.9693	0.9691	0.9719	**0.9982**
	LRSSL	0.7860	0.9212	0.9434	0.7962	0.9462	0.9533	0.9549	0.9693	**0.9964**
Recall	Gdataset	0.8350	0.8412	0.8509	0.7679	0.8827	0.9038	0.9034	0.9015	**0.9753**
	Cdataset	0.8301	0.8477	0.8805	0.7642	0.8945	0.8996	0.9115	0.9119	**0.9826**
	LRSSL	0.8006	0.8310	0.8313	0.7675	0.8925	0.8968	0.8920	0.9091	**0.9717**
NDCG	Gdataset	0.9540	0.9875	0.9931	0.9749	0.9921	0.9917	0.9929	0.9939	**0.9992**
	Cdataset	0.9558	0.9913	0.9946	0.9784	0.9941	0.9953	0.9952	0.9956	**0.9997**
	LRSSL	0.9529	0.9872	0.9911	0.9638	0.9934	0.9923	0.9930	0.9953	**0.9994**
Accuracy	Gdataset	0.7066	0.8404	0.8986	0.7613	0.8856	0.8874	0.8959	0.9077	**0.9688**
	Cdataset	0.7202	0.8701	0.9116	0.7768	0.9001	0.9091	0.9113	0.9154	**0.9821**
	LRSSL	0.7023	0.8385	0.8740	0.6797	0.8896	0.8871	0.8881	0.9127	**0.9711**

TABLE III
ABLATION EXPERIMENTS OF BIBLDR ON GDATASET. **PSC** REPRESENTS PROTOTYPE SPACE CONSTRUCTION. **R-S** AND **D-S** REPRESENTS DRUG AND DISEASE BEHAVIOR SEQUENCE DATA, RESPECTIVELY. **SF** REPRESENTS SIMILARITY FUSION IN STAGE II. **MA** REPRESENTS THE APPLICATION OF MULTI-HEAD SELF-ATTENTION.

	Ablation Components					Gdataset		Cdataset		LRSSL	
	PSC	R-S	D-S	SF	MA	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
[A]	✗	✓	✓	✓	✓	0.9597	0.9678	0.9766	0.9815	0.9031	0.9269
[B]	✓	✗	✓	✓	✓	0.9674	0.9724	0.9798	0.9849	0.9673	0.9774
[C]	✓	✓	✗	✓	✓	0.9131	0.9289	0.9537	0.9606	0.8011	0.8328
[D]	✓	✓	✓	✗	✓	0.9645	0.9616	0.9769	0.9781	0.9533	0.9605
[E]	✓	✓	✓	✓	✗	0.9889	0.9925	0.9906	0.9940	0.9903	0.9932
[F]	✓	✓	✓	✓	✓	0.9941	0.9950	0.9978	0.9982	0.9950	0.9964

fusion strategies to further validate the effectiveness of the mechanism adopted in BiBLDR. We design three approaches. SF1 directly multiplies the similarity values between entities by the prototypes in the behavior sequence rather than using a neural network for feature fusion. SF2 stretches the similarity values into vectors and fuses them with the Transformer layer inputs through neural networks rather than with the prototypes. SF3 stretches the similarity values into vectors and fuses them with the prototypes in behavior sequences through neural networks. This is the strategy adopted in BiBLDR. The detailed experimental results appear in Table IV. The comparison shows that direct numerical multiplication is far less effective than using neural networks to learn the relationship between similarity and prototype vectors. Fusing similarity directly into the behavior sequence prototypes, rather than the raw input vector of the Transformer, proves more effective. Merging similarity with the prototype vector allows easier learning, while trying to integrate it with inconsistent feature spaces complicates the model. Overall, the results validate the effectiveness of BiBLDR's similarity fusion approach.

TABLE IV
THE IMPACT OF DIFFERENT SIMILARITY FUSION STRATEGIES ON BIBLDR PERFORMANCE.

Fusion Strategy	AUROC			AUPRC		
	Gdataset	Cdataset	LRSSL	Gdataset	Cdataset	LRSSL
BiBLDR-SF1	0.9838	0.9911	0.9733	0.9848	0.9929	0.9804
BiBLDR-SF2	0.9835	0.9830	0.9818	0.9827	0.9753	0.9762
BiBLDR-SF3	0.9941	0.9978	0.9950	0.9950	0.9982	0.9964

E. Cold-start studies

To assess BiBLDR's capability to discover therapeutic indications for novel drugs, we perform a drug cold-start experiment on Gdataset. In this setting, new drugs are defined as those lacking known disease associations. For the $i-$th drug u_i, we mask all its disease associations in $\mathcal{A}$, reserving these as the test set. The remaining associations in $\mathcal{A}$ are the training set. This setup evaluates BiBLDR's ability to predict therapeutic effects of u_i across diseases when no prior associations exist, forcing the model to rely exclusively on disease-side behavioral sequences. We test BiBLDR on Gdataset for indication discovery. As shown in Fig. 4, BiBLDR performs best on both metrics. Notably, it demonstrates a significant and overwhelming improvement in AUPRC, reaching 0.7639, while the second-best result is only 0.3484. The AUROC also reaches an impressive 0.9720. These cold-start experiments fully demonstrate the remarkable application value of BiBLDR

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3628673

in novel drug discovery.

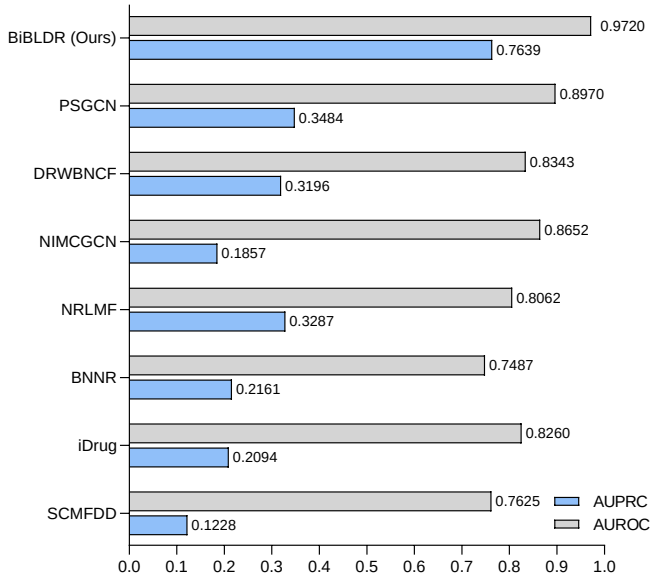


Fig. 4. Comparison of cold-start experimental performance of different methods on the Gdataset.

To more comprehensively evaluate BiBLDR's cold-start performance, we compare its cold-start capabilities against state-of-the-art graph network models in more generalized scenarios. **(1)** ***Drug cold-start:*** when predicting $\mathcal{A}_{ij}$, the association information for drug u_i is not visible. **(2)** ***Disease cold-start:*** when predicting $\mathcal{A}_{ij}$, the association information for disease v_j is not visible. **(3)** ***Blind start:*** when predicting $\mathcal{A}_{ij}$, the association information for both u_i and v_j is not visible. Since blind start for all elements in the association matrix is extremely time-consuming and labor-intensive, we select an equal number of negative samples to predict with all positive samples in blind start scenarios. Detailed performance metrics are presented in the form of a heatmap in Fig. 5.

In both drug cold-start and disease cold-start scenarios, BiBLDR achieves overwhelming superiority across comprehensive metrics. Its AUPRC scores are 0.7639 and 0.5768 respectively, while the second best model only reaches 0.3157 and 0.2239. NDCG scores are 0.9557 and 0.9074, while the second best model only achieves 0.5432 and 0.4780. Although SAGAN and MDGCN achieve high Recall, the underlying reality revealed by Accuracy and NDCG is that they generate a significant number of false positives. In the blind-start scenario, all models show a steep decline in performance. The higher AUPRC and NDCG scores compared to previous scenarios result from fewer test samples and more balanced categories. Under these circumstances, BiBLDR's performance advantages remain satisfactory. The above experiments demonstrate that BiBLDR maintains significant performance advantages even in more generalized and challenging cold-start scenarios.

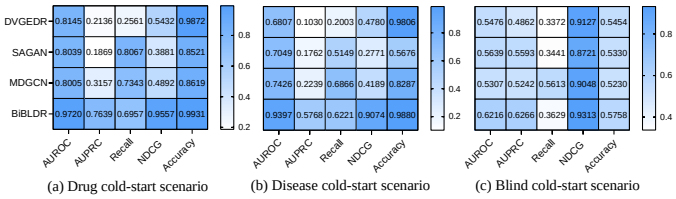


Fig. 5. Comparison of cold-start results across three distinct scenarios.

F. Sparse Environment Testing

Drug repositioning tasks frequently involve managing highly sparse datasets in practical settings. To assess the robustness of BiBLDR under sparse data conditions, we implement a sparse parameter λ, which denotes the fraction of known associations sampled from the original dataset for model training. This configuration establishes a sparse, class-imbalanced data environment. Experiments are conducted by varying lambda across the range of $\{10\%, 20\%, 30\%, 40\%, 50\%, 60\%, 70\%, 80\%, 90\%\}$, and the corresponding results are depicted in the accompanying Fig. 6. The results reveal that even under the extreme sparsity condition of $\lambda = 10\%$, the AUROC values for Gdataset and Cdataset consistently exceed 0.95, while LRSSL achieves a value above 0.94. Furthermore, Gdataset's AUPRC surpasses 0.97, with the other two datasets maintaining values close to 0.97. When known drug-disease associations are extremely sparse, the behavioral sequences derived from the association matrix carry less effective information. BiBLDR still performs robustly because it incorporates drug and disease intrinsic attributes into the association prediction. These prototypes are independent of association sparsity. This experiment shows that using cosine similarity in feature space to model entity similarity is reliable, and semantically rich representations ensure strong performance in sparse scenarios, highlighting BiBLDR's ability to handle limited data.

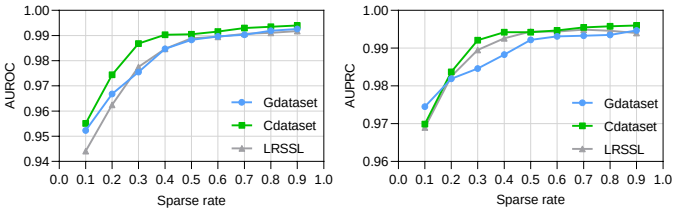


Fig. 6. Sparse environment testing result on Gdataset, Cdataset, and LRSSL.

G. Further Analysis

1) Parameter Analysis: We further investigate the impact of parameters d_0, d_w $\mathcal{T}$, N_{head}, and N_{TL} on the experimental performance. d_0 represents the vector length in the prototype space, sampled from $\{2^6, 2^7, 2^8, 2^9, 2^{10}\}$. d_w represents the vector length of embeddings, sampled from $\{\sqrt{N}, 2^1\sqrt{N}, 2^2\sqrt{N}, 2^3\sqrt{N}\}$. For drug embeddings, N represents the total number of drugs; for disease embeddings, N represents the total number of diseases. $\mathcal{T}$ is the exponential

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3628673

temperature coefficient when integrating the scoring sequence, sampled from $\{1, 2, 3, 4, 5\}$. N_{head} represents the number of attention heads in the Transformer layer, sampled from $\{2^2, 2^3, 2^4, 2^5, 2^6\}$. N_{TL} represents the number of Transformer layer, sampled from $\{1, 2, 3\}$. We conduct experiments using these parameters on the benchmark datasets, and the results are shown in Fig. 7. For d_0, as the values increase, performance shows an upward trend on all datasets, with 2^{10} being the optimal choice. This indicates that higher-dimensional prototypes can represent richer information. For d_w, the majority of studies have already adopted $\sqrt{N}$ as the optimal choice [20], [37], and increasing the embedding length actually leads to a decline in performance, as the original low-dimensional embeddings are already sufficient to capture the drug and disease information during the parameter learning process. An extended length merely introduces redundancy. When $\mathcal{T} > 3$, as the values increase, model performance shows a significant downward trend. This indicates that when the numerical gap between positive and negative samples becomes too large, the model struggles to balance the representation learning of both positive and negative samples. For Gdataset, Cdataset, and LRSSL, the optimal $\mathcal{T}$ parameters are $\{2, 2, 3\}$. For N_{head}, the peak performance is achieved when the Transformer layer contains 8 attention heads. Additionally, experiments demonstrate that increasing the number of Transformer layers leads to a sharp decline in performance, accompanied by a significant increase in computational overhead due to the expansion of model parameters. Excessive Transformer layer parameters complicate model fitting, as additional nonlinear transformations dilute the information. The dataset size is insufficient to support such large-scale parameter adjustments, preventing the model from achieving the desired generalization ability.

2) Visualization Analysis: Cosine similarity constructs a prototype space for drugs and diseases. To validate the rationality of this approach and enhance the interpretability of BiBLDR, we compare the impact of different metric schemes on model performance and visualize the distribution of prototypes in the feature space. In addition to cosine similarity, we select Euclidean distance and Mahalanobis distance as two metrics frequently used for comparing feature vectors. Their calculation formulas are shown in Eq. (12) and Eq. (13), respectively.

$$d_E(\mathbf{A}, \mathbf{B}) = \sqrt{(\mathbf{A} - \mathbf{B})^T (\mathbf{A} - \mathbf{B})} \tag{12}$$

$$d_M(\mathbf{A}, \mathbf{B}) = \sqrt{(\mathbf{A} - \mathbf{B})^T \mathbf{S}^{-1} (\mathbf{A} - \mathbf{B})} \tag{13}$$

where $\mathbf{A}$ and $\mathbf{B}$ denote two feature vectors in the space, and $\mathbf{S}$ is the covariance matrix of the set of feature vectors. The Euclidean distance is the most commonly used distance metric, representing the straight-line distance between two points. In contrast, the Mahalanobis distance extends this concept by incorporating the covariance structure of the data, thereby making it more suitable for analyzing scenarios in which the data exhibits multidimensional distributions. The distribution of prototypes in the feature space, constructed based on three different metric approaches, is shown in Fig. 8. The model performance using different metrics is presented in Table V.

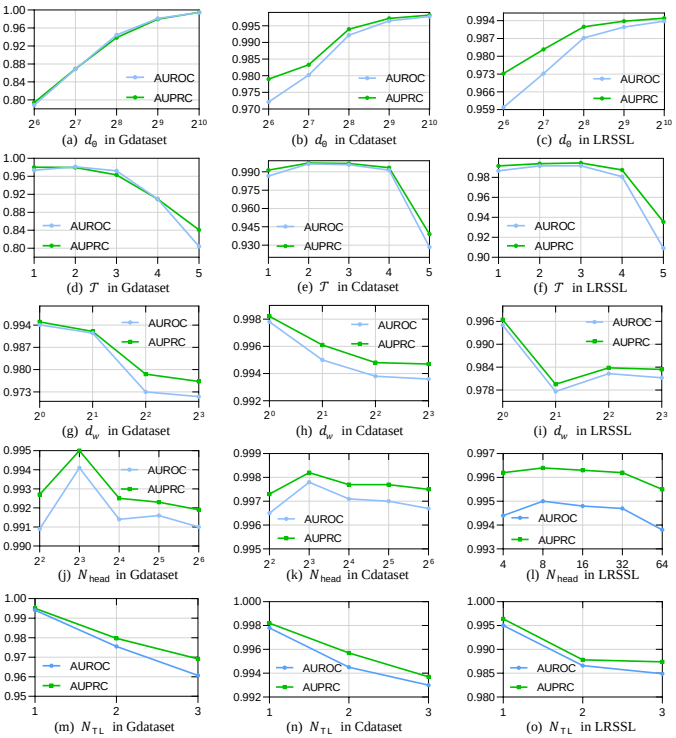


Fig. 7. Parameter analysis experiments on Gdataset, Cdataset, and LRSSL.

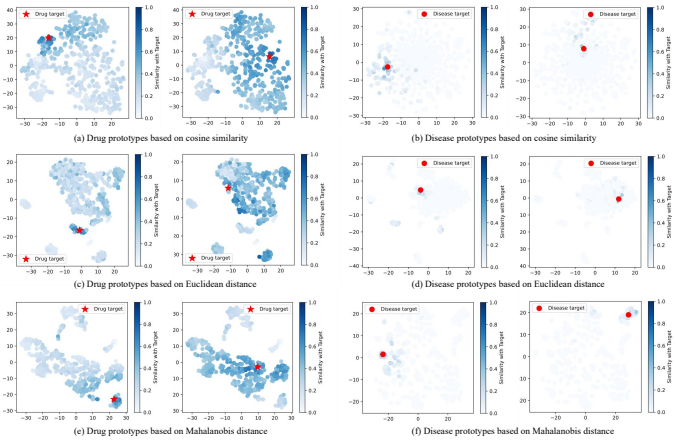


Fig. 8. Visualization of drug and disease prototypes in LRSSL using t-SNE. Darker colored scatter points indicate higher similarity to the target.

TABLE V
MODEL PERFORMANCE UNDER DIFFERENT METRICS. **BIBLDR-E**, **BIBLDR-M**, AND **BIBLDR-C** REPRESENT CONSTRUCTING PROTOTYPE SPACES USING EUCLIDEAN DISTANCE, MAHALANOBIS DISTANCE, AND COSINE SIMILARITY.

Metric	AUROC			AUPRC		
	Gdataset	Cdataset	LRSSL	Gdataset	Cdataset	LRSSL
BiBLDR-E	0.9914	0.9939	0.9944	0.9939	0.9957	0.9962
BiBLDR-M	0.9915	0.9960	0.9947	0.9937	0.9969	0.9964
BiBLDR-C	0.9941	0.9978	0.9950	0.9950	0.9982	0.9964

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3628673

To visualize the distribution of prototypes in feature space, we first perform t-SNE dimensionality reduction on all high dimension drug prototypes and disease prototypes within the dataset and present them in a two-dimensional perspective. Then, entities 11 and 101 (left is No. 11, right is No. 101) are selected as targets in both prototype spaces to observe the similarity distribution of the sample set. Entities with higher similarity to the target in the dataset have darker sample point colors. As shown in Fig. 8, the cosine similarity and Mahalanobis distance demonstrate superior fitting performance compared to the Euclidean distance alone. When analyzing a target containing numerous highly similar samples, the Euclidean distance fails to cluster these closely related samples around the target. Instead, their distribution appears scattered, with many highly similar samples far from the target. In contrast, cosine similarity and Mahalanobis distance effectively cluster the sample points around the target according to their similarity levels, with higher similarity corresponding to closer proximity to the target. Between the two, the distribution of prototypes shaped by cosine similarity in the feature space is more uniform than that produced by the Mahalanobis distance, as cosine similarity places greater emphasis on the directional information of vectors. Moreover, fitting the entire prototype space constitutes a parameter optimization process. In contrast, the Mahalanobis distance, which incorporates the overall covariance of the data, requires a stable covariance matrix. However, during parameter optimization, the variance of the sample population is often unreliable, resulting in covariance estimates that frequently contain substantial noise. This noise adversely affects the fitting performance of the Mahalanobis distance. As shown in Table V, cosine similarity is the optimal choice for constructing prototypes, followed by Mahalanobis distance, while Euclidean distance yields the poorest results.

3) Similarities Analysis: The critical prior information utilized by BiBLDR is drug-drug similarity and disease-disease similarity. Therefore, the manner in which similarity is measured between entities is crucial, as it directly impacts the quality of the generated prototypes and also influences the similarity fusion stage. Therefore, we explore the impact of various drug-drug similarity and disease-disease similarity measures on the performance of BiBLDR. Detailed results are reported in Table VI.

TABLE VI
AN EXPLORATION OF ENTITY SIMILARITY ACROSS DIFFERENT TYPES ON CDATASET. **DR-S** REPRESENTS DRUG-DRUG SIMILARITY TYPE. **DI-S** REPRESENTS DISEASE-DISEASE SIMILARITY TYPE.

Dr-S	Di-S	AUROC	AUPRC	Dr-S	Di-S	AUROC	AUPRC
ChemS	PhS	0.9970	0.9977	ChemS	DoS	0.9961	0.9970
SideS	PhS	0.9969	0.9977	SideS	DoS	0.9956	0.9966
AtcS	PhS	0.9973	0.9976	AtcS	DoS	0.9973	0.9979
DDIS	PhS	0.9971	0.9978	DDIS	DoS	0.9958	0.9969
TargetS	PhS	0.9972	0.9979	TargetS	DoS	0.9963	0.9972

Specifically, we compare five drug-drug similarity measures: *ChemS*, *SideS*, *AtcS*, *DDIS*, *TargetS* and two disease-disease similarity measures: *PhS* and *DoS* on the Cdataset. These multi-similarity data are sourced from [38]. ChemS serves as the drug-drug similarity metric employed by BiBLDR. SideS represents the similarity of drug side effects, with side effects sourced from the SIDER database. The similarity values between drugs are calculated using the Jaccard similarity coefficient. AtcS is ATC Code similarity. The ATC Code is a multi-layer drug classification coding system established by the World Health Organization. Similarity values are obtained through the IC algorithm [39]. DDIS calculates drug interaction profiles using the Jaccard similarity coefficient. TargetS is obtained by calculating the similarity of known targets features of drug using the Jaccard similarity coefficient. All data above are sourced from DrugBank. PhS is the disease-disease similarity metric employed by BiBLDR. Disease ontology can be used to describe relationships between diseases. DoS is calculated using gene ontology-based algorithm [40] to compute disease ontology similarity. Experiments demonstrate that different drug-drug similarities exert a negligible impact on model performance, with differences in both metrics falling below 0.0005. Drug-drug similarity has a greater impact, with overall performance using DoS being approximately 0.001 lower than that of PhS. Overall, the choice of similarity metric between entities does not significantly impact model performance. The prototype representations generated by DoS perform marginally better than those from PhS.

H. Resource Consumption

To validate BiBLDR's practical deployment capabilities and demonstrate its potential value, we evaluate its resource consumption in comparison with several recent models on Gdataset, considering training time, inference time, and parameter number. The detailed results are shown in Table VII. Training time and inference time are calculated based on the average time per fold in 10-fold cross-validation.

TABLE VII
RESOURCE CONSUMPTION COMPARISON EXPERIMENTS ON GDATASET.

Model		Training (s)	Inference (s)	Parameters (Mb)
DVGEDR		148.29	1.47	0.28
SAGAN		823.55	15.16	9.12
MDGCN		720.13	0.12	47.09
BiBLDR	$d_0 = 1024$	14.39	1.37	148.98
BiBLDR	$d_0 = 512$	12.45	1.13	43.92
BiBLDR	$d_0 = 128$	11.55	1.08	6.03

As shown in Table 5, the number of parameters in BiBLDR depends on the prototype length, since longer prototypes lead to a sharp increase in the parameters of the FFN layer within the Transformer layer. To enable the deployment of BiBLDR on mobile devices, the number of parameters is reduced by decreasing the prototype length. Experiments in the parameter analysis section show that the performance after parameter reduction remains acceptable. Although BiBLDR becomes a large-scale model when it uses long prototypes, its training efficiency provides a significant advantage. This advantage arises because BiBLDR leverages rich behavioral sequence data, which enables the model to converge within a few epochs. BiBLDR's low inference time also enables seamless deployment on mobile devices.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3628673

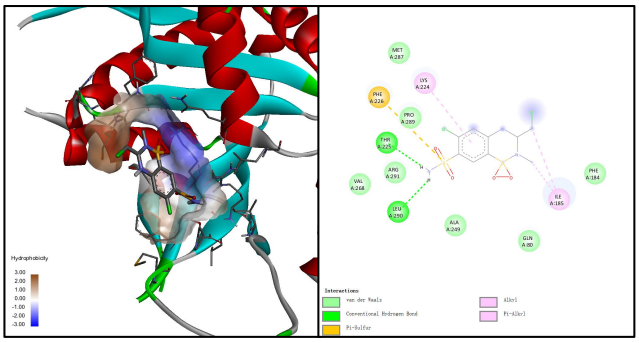
I. Case Studies

To evaluate the real-world applicability of BiBLDR in drug repositioning, we perform case studies focusing on two diseases: lung cancer and hypertension. Specifically, we utilize all known drug-disease associations in Gdataset as the training set to train the model. Subsequently, all rest associations with lung cancer and hypertension are treated as candidate pairs, and the trained model is used to predict scores for each candidate pair. For both lung cancer and hypertension, the top-10 candidate drugs are ranked according to their predicted scores, and each was individually validated to confirm the existence of a therapeutic relationship.

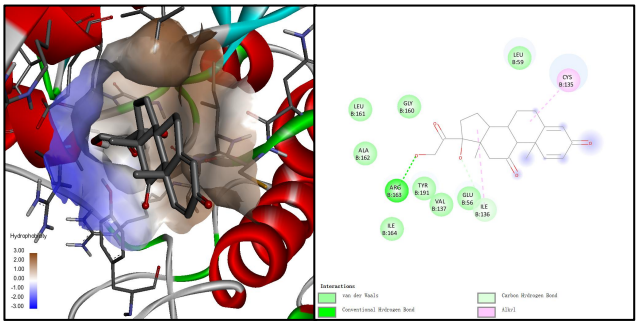
1) Lung Cancer (211980 in OMIM): Lung cancer is one of the cancers with the highest incidence and mortality rates worldwide, feared for its high fatality rate and lack of noticeable early symptoms. The left of Table VIII provides the top-10 predicted candidate drugs for lung cancer, as determined by BiBLDR. Eight of the 10 candidate lung cancer drugs are validated as accurate predictions. For instance, combination chemotherapy regimens involving cisplatin or carboplatin with etoposide have demonstrated effectiveness in treating both non-small and small cell lung cancer [41]. Etoposide is the highest-ranked candidate drug in Table VIII.

2) Hypertension (145500 in OMIM): Hypertension is a prevalent and highly detrimental chronic condition, known for its subtle onset and multifaceted causes. It is especially common among middle-aged and elderly individuals, though it can also occur in younger populations. The right of Table VIII provides the top-10 predicted candidate drugs for hypertension. Nine of the top-ranked candidate drugs predicted by BiBLDR have been supported by valid evidence, achieving a 90% hit rate. For instance, methyclothiazide has been proven effective in treating hypertension caused by heart failure, kidney failure, or estrogen and steroid therapy. Methyclothiazide is the tenth-ranked candidate drug in Table VIII.

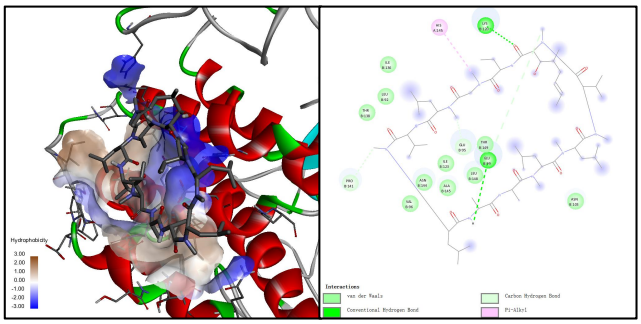
3) Molecular Docking Experiment: For the unconfirmed drug-disease associations listed in Table VIII, molecular docking experiments are performed using Autodock Vina and Discovery Studio software to investigate the interaction between drug ligands and target proteins. The binding free energies were calculated to evaluate the pharmacological relevance. The specific molecular docking experiment is illustrated in Fig. 9. In the prediction of the association between Methyclothiazide and lung cancer, BCCIP (PDB Code: 7KYQ) is selected as the target protein. Fig. 9(a) illustrates the molecular interactions between the hydrochlorothiazide ligand and specific amino acid residues of BCCIP (MET287, PRO289, ARG291, VAL268, ALA249, GLN80, and PHE184), including van der Waals forces, conventional hydrogen bonds between nitrogen atoms and residues (THR255, LEU290), and a $\pi-$sulfur interaction (PHE226). The ligand-protein complex exhibited a binding free energy of -11.25 kcal/mol. In the prediction of the association between Prednisone and lung cancer, CDK4 (PDB Code: 2W9Z) is selected as the target protein. Fig. 9(b) illustrates the van der Waals interactions between the prednisone ligand and specific amino acid residues of CDK4 (GLY160, LEU161, ALA162, ILE164, TYR191, VAL137, GLU56, and LEU59), a conventional hydrogen bond formed between an oxygen atom and residues (ARG163), and a carbon-hydrogen bond (ILE136). The ligand-protein complex exhibited a binding free energy of -14.9397 kcal/mol. In the prediction of the association between Cyclosporin and hypertension, the Hexameric (PDB Code: 1P9M) is selected as the target protein. Fig. 9(c) illustrates the van der Waals interactions between the cyclosporine ligand and specific amino acid residues of the Hexameric (ILE136, LEU92, THR138, THR149, ILE123, ASN144, ALA145, VAL96, and ASN103), conventional hydrogen bonds formed between oxygen and nitrogen atoms and the residues (LYS120 and GLU99), and a carbon-hydrogen bond (GLU95). Although some of the potentially effective drug-disease associations predicted by BibLDR have not yet been validated by existing evidence, molecular docking experiments and the calculated binding free energies demonstrate the potential value of these predicted associations by BiBLDR.



(a) Interactions of Methyclothiazide with (PDB code: 7KYQ)



(b) Interactions of Prednisone with (PDB code: 2W9Z)



(c) Interactions of Cyclosporine with (PDB code: 1P9M)

Fig. 9. The molecular docking results of ligand molecules with target proteins.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3628673

TABLE VIII
TOP-10 CANDIDATE DRUG OF LUNG CANCER AND HYPERTENSION. LEFT IS LUNG CANCER AND RIGHT IS HYPERTENSION.

Rank	Lung Cancer			Hypertension		
	DrugBank ID	Candidate Drug	Evident	DrugBank ID	Candidate Drug	Evident
1	DB00773	Etoposide	[41]	DB01023	Felodipine	[42]
2	DB01030	Topotecan	[43]	DB01244	Bepridil	[44]
3	DB00563	Methotrexate	[45]	DB00310	Chlorthalidone	[46]
4	DB00232	Methyclothiazide	-	DB01611	Hydroxychloroquine	[47]
5	DB00762	Irinotecan	[48]	DB00808	Indapamide	[49]
6	DB00853	Temozolomide	[50]	DB00800	Fenoldopam	[51]
7	DB01268	Sunitinib	[52]	DB00091	Cyclosporine	-
8	DB00694	Daunorubicin	[53]	DB00178	Ramipril	[54]
9	DB00859	Penicillamine	[55]	DB00880	Chlorothiazide	[56]
10	DB00635	Prednisone	-	DB00232	Methyclothiazide	[57]

IV. CONCLUSION

In this paper, we propose a bidirectional behavior learning strategy for drug repositioning. Unlike previous graph representation-based deep learning methods, we frame the drug repositioning task as a recommendation system problem, leveraging behavioral sequence analysis to predict drug-disease associations. We present a novel bidirectional behavioral sequence construction method that simultaneously considers the behavioral sequences of both drug and disease entities, enriching the semantic representation of these sequences. Additionally, we incorporate the Transformer architecture to further model the interactions between different behavioral sequences. Extensive experiments demonstrate that our proposed BiBLDR achieves state-of-the-art performance. The various extension experiments provide solid support for the interpretability of BiBLDR, while also validating its valuable potential in drug repositioning.

In the future, we aim to capture global similarity during prototype construction to shape entity prototypes, moving beyond pairwise relationships. We also plan to enhance prototype representation by incorporating richer semantic information. Additionally, explicit fusion strategies for similarity and integration of drug-disease association scores will be explored for bidirectional behavioral sequences. Moreover, more comprehensive experiments can be conducted to examine how different attention mechanisms influence the processing of bidirectional behavioral sequences. Meanwhile, some emerging GNN technologies can also be considered for integration into the BiBLDR framework [58], [59]. Compared with simple similarity matrices, GNNs can model topological relationships between entities and extract effective spatial semantic information. With these advantages, GNNs can be further introduced into the process of modelling similarity information and extracting similarity features.

REFERENCES

[1] K. Huang, P. Chandak, Q. Wang, S. Havaldar, A. Vaid, J. Leskovec, G. N. Nadkarni, B. S. Glicksberg, N. Gehlenborg, and M. Zitnik, "A foundation model for clinician-centered drug repurposing," *Nature Medicine*, pp. 1–13, 2024.

[2] M. A. Farha and E. D. Brown, "Drug repurposing for antimicrobial discovery," *Nature microbiology*, vol. 4, no. 4, pp. 565–577, 2019.

[3] S. Ning, "Research progress on the application of aspirin and its derivatives in the treatment and rehabilitation of cardiovascular diseases," in *E3S Web of Conferences*, vol. 553. EDP Sciences, 2024, p. 04010.

[4] D. Zhang, C. LaSenna, and B. E. Shields, "Culprits of medication-induced telogen effluvium, part 2." *Cutis*, vol. 113, no. 1, pp. 11–14, 2024.

[5] R. L. Gottlieb and R. Paredes, "Oral and intravenous 1'-cyano-substituted adenosine-like antivirals for early covid-19," *The Lancet Infectious Diseases*, vol. 24, no. 2, pp. 108–110, 2024.

[6] M. Yang, H. Luo, Y. Li, and J. Wang, "Drug repositioning based on bounded nuclear norm regularization," *Bioinformatics*, vol. 35, no. 14, pp. i455–i463, 2019.

[7] X. Su, P. Hu, H. Yi, Z. You, and L. Hu, "Predicting drug-target interactions over heterogeneous information network," *IEEE journal of biomedical and health informatics*, vol. 27, no. 1, pp. 562–572, 2022.

[8] B.-W. Zhao, X.-R. Su, Y. Yang, D.-X. Li, G.-D. Li, P.-W. Hu, Z.-H. You, X. Luo, and L. Hu, "Regulation-aware graph learning for drug repositioning over heterogeneous biological network," *Information Sciences*, vol. 686, p. 121360, 2025.

[9] B.-W. Zhao, X.-R. Su, P.-W. Hu, Y.-P. Ma, X. Zhou, and L. Hu, "A geometric deep learning framework for drug repositioning over heterogeneous information networks," *Briefings in Bioinformatics*, vol. 23, no. 6, p. bbac384, 2022.

[10] Y. Meng, C. Lu, M. Jin, J. Xu, X. Zeng, and J. Yang, "A weighted bilinear neural collaborative filtering approach for drug repositioning," *Briefings in bioinformatics*, vol. 23, no. 2, p. bbab581, 2022.

[11] X. Sun, B. Wang, J. Zhang, and M. Li, "Partner-specific drug repositioning approach based on graph convolutional network," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 11, pp. 5757–5765, 2022.

[12] B.-W. Zhao, X.-R. Su, P.-W. Hu, Y.-A. Huang, Z.-H. You, and L. Hu, "igrldti: an improved graph representation learning method for predicting drug–target interactions over heterogeneous biological information network," *Bioinformatics*, vol. 39, no. 8, p. btad451, 2023.

[13] X. Meng, X. Wen, X. Sun, J. Xiang, Y. Long, X. Lin, and M. Li, "Sagan: A subgraph-aware graph attention network for drug repositioning," *Tsinghua Science and Technology*, 2025.

[14] H. Wang, Q. Gu, J. Wei, Z. Cao, and Q. Liu, "Mining drug–disease relationships as a complement to medical genetics-based drug repositioning: Where a recommendation system meets genome-wide association studies," *Clinical Pharmacology & Therapeutics*, vol. 97, no. 5, pp. 451–454, 2015.

[15] M. G. Ozsoy, T. Özyer, F. Polat, and R. Alhajj, "Realizing drug repositioning by adapting a recommendation system to handle the process," *BMC bioinformatics*, vol. 19, pp. 1–14, 2018.

[16] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *ACM computing surveys (CSUR)*, vol. 52, no. 1, pp. 1–38, 2019.

[17] A. Shankar, P. Perumal, M. Subramanian, N. Ramu, D. Natesan, V. R. Kulkarni, and T. Stephan, "An intelligent recommendation system in e-commerce using ensemble learning," *Multimedia Tools and Applications*, vol. 83, no. 16, pp. 48 521–48 537, 2024.

[18] K. Sharma, Y.-C. Lee, S. Nambi, A. Salian, S. Shah, S.-W. Kim, and

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3628673

S. Kumar, "A survey of graph neural networks for social recommender systems," *ACM Computing Surveys*, vol. 56, no. 10, pp. 1–34, 2024.

[19] Y. Li, C. P. Chen, and T. Zhang, "A survey on siamese network: Methodologies, applications, and opportunities," *IEEE Transactions on artificial intelligence*, vol. 3, no. 6, pp. 994–1014, 2022.

[20] Q. Chen, H. Zhao, W. Li, P. Huang, and W. Ou, "Behavior sequence transformer for e-commerce recommendation in alibaba," in *Proceedings of the 1st international workshop on deep learning practice for high-dimensional sparse data*, 2019, pp. 1–4.

[21] A. Gottlieb, G. Y. Stein, E. Ruppin, and R. Sharan, "Predict: a method for inferring novel drug indications with application to personalized medicine," *Molecular systems biology*, vol. 7, no. 1, p. 496, 2011.

[22] H. Luo, J. Wang, M. Li, J. Luo, X. Peng, F.-X. Wu, and Y. Pan, "Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm," *Bioinformatics*, vol. 32, no. 17, pp. 2664–2671, 2016.

[23] X. Liang, P. Zhang, L. Yan, Y. Fu, F. Peng, L. Qu, M. Shao, Y. Chen, and Z. Chen, "Lrssl: predict and interpret drug–disease associations based on data integration using sparse subspace learning," *Bioinformatics*, vol. 33, no. 8, pp. 1187–1196, 2017.

[24] D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, and J. Woolsey, "Drugbank: a comprehensive resource for in silico drug discovery and exploration," *Nucleic acids research*, vol. 34, no. suppl_1, pp. D668–D672, 2006.

[25] A. Hamosh, A. F. Scott, J. S. Amberger, C. A. Bocchini, and V. A. McKusick, "Online mendelian inheritance in man (omim), a knowledge-base of human genes and genetic disorders," *Nucleic acids research*, vol. 33, no. suppl_1, pp. D514–D517, 2005.

[26] A. P. Davis, C. J. Grondin, R. J. Johnson, D. Sciaky, B. L. King, R. McMorran, J. Wiegers, T. C. Wiegers, and C. J. Mattingly, "The comparative toxicogenomics database: update 2017," *Nucleic acids research*, vol. 45, no. D1, pp. D972–D978, 2017.

[27] X. Sun, X. Jia, Z. Lu, J. Tang, and M. Li, "Drug repositioning with adaptive graph convolutional networks," *Bioinformatics*, vol. 40, no. 1, p. btad748, 2024.

[28] C. Steinbeck, C. Hoppe, S. Kuhn, M. Floris, R. Guha, and E. L. Willighagen, "Recent developments of the chemistry development kit (cdk)-an open-source java library for chemo-and bioinformatics," *Current pharmaceutical design*, vol. 12, no. 17, pp. 2111–2120, 2006.

[29] D. Rogers and M. Hahn, "Extended-connectivity fingerprints," *Journal of chemical information and modeling*, vol. 50, no. 5, pp. 742–754, 2010.

[30] C. Cavestro, A. Rosatello, M. P. Marino, G. Micca, and G. Asteggiano, "High prolactin levels as a worsening factor for migraine," *The Journal of Headache and Pain*, vol. 7, no. 2, pp. 83–89, 2006.

[31] C. E. Lipscomb, "Medical subject headings (mesh)," *Bulletin of the Medical Library Association*, vol. 88, no. 3, p. 265, 2000.

[32] W. Zhang, X. Yue, W. Lin, W. Wu, R. Liu, F. Huang, and F. Liu, "Predicting drug-disease associations by using similarity constrained matrix factorization," *BMC bioinformatics*, vol. 19, pp. 1–12, 2018.

[33] H. Chen, F. Cheng, and J. Li, "idrug: Integration of drug repositioning and drug-target prediction via cross-network embedding," *PLoS computational biology*, vol. 16, no. 7, p. e1008040, 2020.

[34] J. Li, S. Zhang, T. Liu, C. Ning, Z. Zhang, and W. Zhou, "Neural inductive matrix completion with graph convolutional networks for mirna-disease association prediction," *Bioinformatics*, vol. 36, no. 8, pp. 2538–2546, 2020.

[35] D. Niu, L. Zhang, B. Zhang, Q. Zhang, S. Ding, H. Wei, and Z. Li, "Dvgedr: a drug repositioning method based on dual-view fusion and graph enhancement mechanism in heterogeneous networks," *Complex & Intelligent Systems*, vol. 11, no. 1, p. 68, 2025.

[36] Y. Gan, S. Li, G. Xu, C. Yan, and G. Zou, "Multidependency graph convolutional networks and contrastive learning for drug repositioning," *Journal of Chemical Information and Modeling*, vol. 65, no. 6, pp. 3090–3103, 2025.

[37] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.

[38] M. Yang, G. Wu, Q. Zhao, Y. Li, and J. Wang, "Computational drug repositioning based on multi-similarities bilinear matrix factorization," *Briefings in bioinformatics*, vol. 22, no. 4, p. bbaa267, 2021.

[39] P. Resnik, "Using information content to evaluate semantic similarity in a taxonomy," in *Proceedings of the 14th international joint conference on Artificial intelligence-Volume 1*, 1995, pp. 448–453.

[40] J. Z. Wang, Z. Du, R. Payattakool, P. S. Yu, and C.-F. Chen, "A new method to measure the semantic similarity of go terms," *Bioinformatics*, vol. 23, no. 10, pp. 1274–1281, 2007.

[41] J. Y. Jang, D. Kim, E. Im, and N. D. Kim, "Etoposide as a key therapeutic agent in lung cancer: Mechanisms, efficacy, and emerging strategies," *International Journal of Molecular Sciences*, vol. 26, no. 2, p. 796, 2025.

[42] P. Mace, T. Stallard, and W. Littler, "Felodipine in hypertension," *European journal of clinical pharmacology*, vol. 29, pp. 383–389, 1985.

[43] N. Horita, M. Yamamoto, T. Sato, T. Tsukahara, H. Nagakura, K. Tashiro, Y. Shibata, H. Watanabe, K. Nagai, M. Inoue *et al.*, "Topotecan for relapsed small-cell lung cancer: systematic review and meta-analysis of 1347 patients," *Scientific reports*, vol. 5, no. 1, p. 15437, 2015.

[44] K. AOKI and K. SATO, "Hypotensive effects of a calcium antagonist, bepridil, in patients with essential hypertension," *Japanese heart journal*, vol. 27, no. 5, pp. 693–699, 1986.

[45] J. Smyth and H. Ford, "Methotrexate in the chemotherapy of lung cancer." *Cancer Treatment Reports*, vol. 65, pp. 161–163, 1981.

[46] D. B. Toubes, T. J. McIntosh, W. M. Kirkendall, and W. R. Wilson, "Hypotensive effects of clonidine and chlorthalidone: Controlled clinical trial of drugs administered singly and in combination," *American Heart Journal*, vol. 82, no. 3, pp. 312–318, 1971.

[47] M. Choi, J. R. Hwang, J.-H. Sung, N. Byun, Y. S. Seok, G. J. Cho, S.-J. Choi, J.-S. Kim, S.-y. Oh, and C.-R. Roh, "Hydroxychloroquine reduces hypertension and soluble fms-like kinase-1 in a nω-nitro-l-arginine methyl ester-induced preeclampsia rat model," *Journal of Hypertension*, vol. 40, no. 12, pp. 2459–2468, 2022.

[48] A. Sevinc, M. E. Kalender, M. Altinbas, M. Ozkan, M. Dikilitas, C. Camci, A. S. of Medical Oncology (ASMO) *et al.*, "Irinotecan as a second-line monotherapy for small cell lung cancer," *Asian Pac J Cancer Prev*, vol. 12, no. 4, pp. 1055–1059, 2011.

[49] N. M. Kaplan, "Indapamide: is it the better diuretic for hypertension?" pp. 983–984, 2015.

[50] R. Dziadziuszko, A. Ardizzoni, P. Postmus, E. Smit, A. Price, C. Debruyne, C. Legrand, G. Giaccone, E. L. C. Group *et al.*, "Temozolomide in patients with advanced non-small cell lung cancer with and without brain metastases: a phase ii study of the eortc lung cancer group (08965)," *European Journal of Cancer*, vol. 39, no. 9, pp. 1271–1276, 2003.

[51] M. B. Murphy, C. Murray, and G. D. Shorten, "Fenoldopam—a selective peripheral dopamine-receptor agonist for the treatment of severe hypertension," *New England Journal of Medicine*, vol. 345, no. 21, pp. 1548–1557, 2001.

[52] S. Tanday, "Sunitinib benefits patients with small-cell lung cancer," *The Lancet Oncology*, vol. 16, no. 4, p. e164, 2015.

[53] Y. Wang, M. Fu, J. Liu, Y. Yang, Y. Yu, J. Li, W. Pan, L. Fan, G. Li, X. Li *et al.*, "Inhibition of tumor metastasis by targeted daunorubicin and dioscin codelivery liposomes modified with pfv for the treatment of non-small-cell lung cancer," *International journal of nanomedicine*, pp. 4071–4090, 2019.

[54] J. E. Frampton and D. H. Peters, "Ramipril: an updated review of its therapeutic use in essential hypertension and heart failure," *Drugs*, vol. 49, pp. 440–466, 1995.

[55] S. J. Sciegienka, S. R. Solst, K. C. Falls, J. D. Schoenfeld, A. R. Klinger, N. L. Ross, S. N. Rodman, D. R. Spitz, and M. A. Fath, "D-penicillamine combined with inhibitors of hydroperoxide metabolism enhances lung and breast cancer cell responses to radiation and carboplatin via h2o2-mediated oxidative stress," *Free Radical Biology and Medicine*, vol. 108, pp. 354–361, 2017.

[56] E. D. Freis, A. Wanko, I. M. Wilson, A. E. Parrish *et al.*, "Chlorothiazide in hypertensive and normotensive patients," *Ann NY Acad Sci*, vol. 71, no. 4, pp. 450–455, 1958.

[57] H. R. Black, S. G. Chrysant, C. L. Curry, W. H. Frishman, R. H. Grimm, K. C. Lasseter, R. Okun, J. L. Pool, V. Raizada, N. D. Vlachakis *et al.*, "Antihypertensive and metabolic effects of concomitant administration of terazosin and methyclothiazide for the treatment of essential hypertension," *The Journal of Clinical Pharmacology*, vol. 32, no. 4, pp. 351–359, 1992.

[58] Y. Yang, G. Li, D. Li, J. Zhang, P. Hu, and L. Hu, "Integrating fuzzy clustering and graph convolution network to accurately identify clusters from attributed graph," *IEEE Transactions on Network Science and Engineering*, 2024.

[59] Y. Yang, L. Hu, G. Li, D. Li, P. Hu, and X. Luo, "Link-based attributed graph clustering via approximate generative bayesian learning," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2025.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/