R-DTI: Drug Target Interaction Prediction Based on Second-Order Relevance Exploration

Yang Hua[1,2,3], Tianyang Xu[1,2,3], Xiaoning Song[1,2,3*], Zhenhua Feng[1,2,3], Rui Wang[1,2,3], Wenjie Zhang[1,2,3], Xiaojun Wu[1,2,3]

[1]School of Artifcial Intelligence and Computer Science, Jiangnan University, Wuxi, Jiangsu, P.R. China
[2]Sino-UK Joint Laboratory on Artificial Intelligence, Ministry of Science and Technology, China
[3]International Joint Laboratory on Artificial Intelligence, Ministry of Education, China
{7211905018; wenjie.zhang}@stu.jiangnan.edu.cn; {tianyang.xu; x.song; fengzhenhua; cs_wr; wu_xiaojun}@jiangnan.edu.cn;

Abstract

Drug Target Interaction (DTI) prediction has witnessed promising performance boosts accompanied by advanced multimodal feature extraction. However, existing approaches suffer from two main difficulties. First, the complex protein structures cannot be well represented by current protein-sequence-based feature extractors. Second, the gap between protein and drug features increases the vulnerability of the obtained classifier thus degrading the prediction robustness. To address these issues, we propose a novel R-DTI method by exploring the second-order relevance in both protein structural feature extraction and DTI prediction phases. Specifically, we construct a pre-trained structural feature extractor that mines the atomic relevance of each amino acid. Then, an inter-feature structure-preserved Riemannian network is designed to expand the existing protein extraction patterns. To improve the prediction robustness, we also develop a Riemannian classifier that uses the second-order protein-drug relevance with a unified feature space. Extensive experimental results prove the merits and superiority of our R-DTI against the state-of-the-art, achieving 1.4% and 1.9% higher AUC-ROC on the BindingDB and DrugBank datasets, respectively.

Code — https://github.com/JU-HuaY/R-DTI

Introduction

Drug Target Interaction (DTI) prediction is critical to virtual drug screening and accelerating drug discovery pipelines (Cao et al. 2012; Coelho, Arrais, and Oliveira 2016). The DTI prediction methods are mainly designed for two tasks, classification (Jacob and Vert 2008) and regression (Tang et al. 2014), which aim to qualify and quantify DTI, respectively. Since both tasks rely on similar backbones, deep-learning-based DTI prediction models (Öztürk, Özgür, and Ozkirimli 2018; Lee, Keum, and Nam 2019) excel at both tasks due to their large parameter capacity. In general, their superior performance depends on two key factors: feature extraction and prediction formulation.

For feature extraction, deep-learning-based DTI prediction models often extract protein and drug features from the textual and structural modalities (Hua et al. 2023a, 2025). The recent development of textual feature extraction is rapid, particularly with the advent of Large Language Models (LLMs) (Elnaggar et al. 2021; Wang et al. 2019), which can generate key semantic features based on pre-trained models. For drug structural features, many attempts (Hua et al. 2023a,b) can directly transform the Simplified Molecular-Input Line-Entry System (SMILES) equations into graph structure features by the molecular analysis tool, Rdkit (Landrum et al. 2016). However, the complexity of protein structures presents challenges in their representation. As shown in Figure 1 (a), Lim et al. (2019) proposed a graph attention algorithm to predict DTI from data with 3D protein structures. However, the method only considers a small number of samples to construct protein structural information, which lacks generalization for diverse proteins. Wu et al. (2024) proposes a pre-trained encoder that first addresses the inability to extract structural features without structure data. The encoder is trained by contrastive learning based on datasets containing contact matrices of amino acids, as shown in Figure 1 (b). However, without the 3D properties and corresponding topology, the contact matrices are insufficient to represent protein structural features (Hua et al. 2024). Therefore, how to improve the quality of the extracted protein structural features remains to be explored.

For prediction formulation, existing methods mainly rely on the analysis of Euclidean DTI representations, including two types, drug-target joint features (**first-order relevance**) and drug-target relationship features (**second-order relevance**). Both types suffer from typical limitations. As shown in Figure 1, the first-order relevance prediction methods (Lee, Keum, and Nam 2019; Abbasi et al. 2020) incorporate data-driven influences that bias the classifier training toward protein features, while suppressing the drug clues. While the DTI methods (Chen et al. 2020; Wu et al. 2024) based on second-order relevance prediction can effectively address this issue, they still have two problems. 1) The relevance calculation involves matrix multiplication, which confuses the semantic information of the DTI feature if the protein and drug feature spaces are not consistently unified. 2) The second-order relevance is unstable due to its excessive information density, challenging the convergence state.

To address these issues, we propose a novel R-DTI framework based on second-order relevance exploration. For pro-

*Xiaoning Song is the corresponding author.

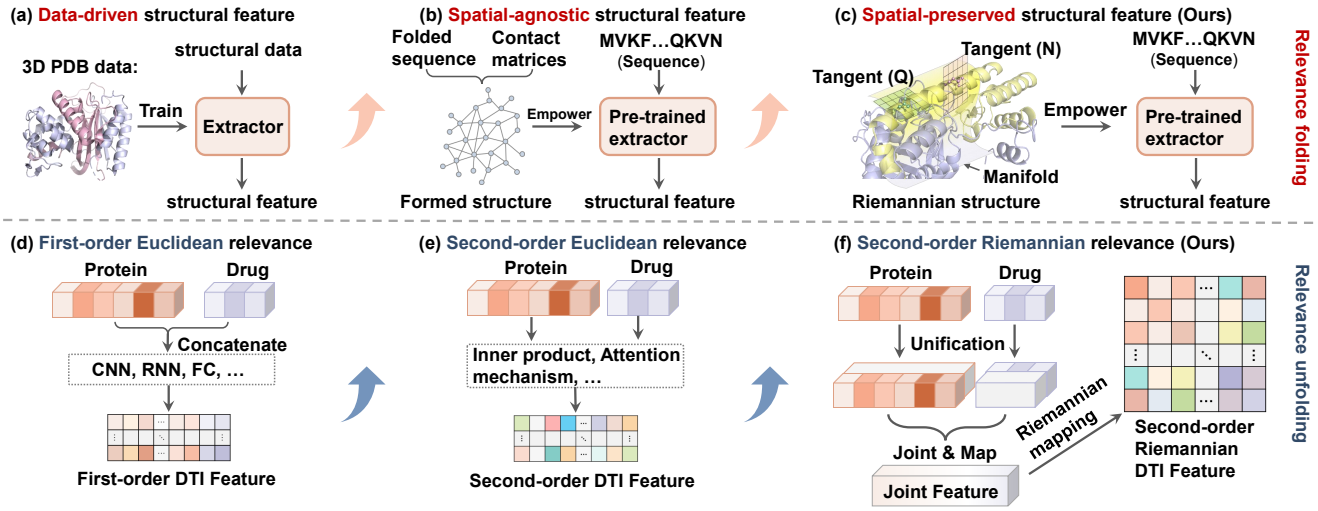

Figure 1: (a), (b) and (c) show three structural feature extractors. (d) and (e) show the first- and second-order Euclidean relevance modeling for DTI prediction. (f) shows our proposed second-order Riemannian relevance modeling for DTI prediction.

tein structural feature extraction, we design an extractor according to the structural properties of amino acids accompanied by the Riemannian space metric. Specifically, we construct the covariance matrix of atomic vectors to emphasize second-order relevance within amino acids. Meanwhile, to preserve the topological relationships between amino acids within a protein, we consider amino acids distributed on the protein manifold. Mathematically, the constructed covariance matrix is precisely lying on the corresponding tangent space, which describes the second-order Riemannian relevance of constituent atoms (Arsigny et al. 2007), as shown in Figure 1 (c). Given this consideration, we construct a pre-trained sequence-based extractor that learns enough priori structural knowledge from the given protein structure. For prediction formulation, to compensate for the learning bias of the DTI features and stabilize the learning stage, we extract the second-order Riemannian relevance of the DTI features for DTI prediction. We first unify the protein and drug feature spaces, and map the concatenated protein and drug features into second-order Riemannian DTI features. Then, we project the DTI features to a lower dimensional space to ensure their stability. Last, the stable DTI features are used to achieve the final prediction by a linear classifier.

We evaluate the proposed method on six benchmarking datasets. Extensive experimental results show that our R-DTI achieves superior performance against the state-of-the-art. Furthermore, our visualization investigation supports the merit of exploiting second-order relevance matrices in highlighting the focal regions of biological molecules.

In summary, the main contributions of R-DTI include:

- A novel R-DTI framework that introduces the Riemannian manifold to DTI prediction for the first time and outperforms the state-of-the-art on six public datasets.
- A new sequence-based protein structural feature extractor that contains priori atomic properties and protein topological information.
- A Riemannian classifier that compensates for the bias of the DTI features, stabilizing the performance.

Related Work

Drug Target Interaction Advances

DTI prediction is a task that discriminates whether there is an interaction between a protein and a drug based on their features. The prediction algorithms usually contain three processes: protein feature extraction, drug feature extraction, and interaction prediction (Hua et al. 2023b). In the primary research stage, Yamanishi et al. (2008) used handcrafted features for protein and drug feature engineering, followed by shallow machine learning solutions for DTI prediction. Traditional methods are based on rigorous theoretical derivation, resulting in excellent interpretability of both the representation and the prediction analysis. However, shallow methods are limited by their model capacity and hardware constraints.

With the development of computing devices and deep learning methods (Song et al. 2024), many advanced DTI prediction methods, such as DeepDTA (Öztürk, Özgür, and Ozkirimli 2018) and HADTI (Zhao et al. 2022), have attracted wide attention. These methods demonstrate the potential to achieve high-precision DTI prediction in constructing models with a large number of parameters. However, these deep methods are in a black-box state for both feature extraction and decision making. Therefore, the lack of understandable theoretical evidence complicates the explanation of their advantages.

Riemannian Network Techniques

Deep neural networks have demonstrated their advantages in many AI tasks (Xu, Zhu, and Wu 2023; Tang, Wang, and Hu 2023). The hierarchically obtained feature maps in the dense Euclidean space are the most essential components. However, practical data features defined by coordinates adhere to

non-Euclidean geometries (Huang and Van Gool 2017). To tackle these non-Euclidean features, Helgason (1965) first proposed a notion of local geodesic convolution on non-Euclidean domains to extract local patches on the shape manifold. On this basis, Huang and Van Gool (2017) developed a symmetric positive definite (SPD) manifold and proposed the SPDNet, which comprises BiMap and ReEig. These components are inspired by the Euclidean linear and ReLU layers, respectively.

Given the intrinsic low-dimension modelling power, the Riemannian SPD manifold has been successfully applied in the biological field. To learn a better representation of protein, Detlefsen, Hauberg, and Boomsma (2022) developes a suitable Riemannian metric that corresponds to geodesic distances between one-hot encoded proteins integrated along the manifold. Besides, Wang et al. (2022) proposes a harmonic molecular representation learning framework that provides a multi-resolution representation of molecular geometric and chemical features on a 2D Riemannian manifold. These studies investigate the feasibility of modeling networks on the Riemannian manifold for bioinformatics.

The Proposed R-DTI Method

The proposed R-DTI contains three components, *i.e.*, two feature extractors to encode the protein and drug data, respectively, and the Riemannian classifier for final prediction.

Multi-Mode Protein Feature Extractor

We use pre-trained extractors to extract textual and structural features from protein sequences, as shown in Figure 2. First, Prot-Bert (Elnaggar et al. 2021), a Large-Language-Model (LLM), and DenseNet (Hua et al. 2023b) are used to sequentially represent and extract protein textual features, $\mathbf{F}_p^t \in \mathbb{R}^{L_p \times C_t}$, where L_p and C_t are the sequence length and the textual feature channel, respectively. Then, we pre-train a sequence-based extractor that could complement the protein structural information. Specifically, we use a structure-based extractor and 3D PDB data (Berman et al. 2000) containing protein sequence and validated 3D structure to empower the sequence-based extractor. To align the 3D structure with the sequence, we rewrite the 3D data from atom sets, $\{(\mathbf{p}_i^a, \mathbf{c}_i^a)\}_{i=1}^N$, to amino acid sets, $\{(\mathbf{P}_i, \mathbf{c}_i)\}_{i=1}^K$, where $\mathbf{p}_i^a$ and $\mathbf{c}_i^a$ are the property vector of the i-th atom and its coordinate, $\mathbf{P}_i$ and $\mathbf{c}_i$ are the property matrix of the i-th amino acid and its coordinate, N and K are the atom and amino acid numbers, respectively. Specifically, the covariance of atomic property vectors, $\{\mathbf{p}_0^a, ..., \mathbf{p}_{N_i}^a\} \in \mathbb{R}^{C_s \times N_i}$, is used as the amino acid property, $\mathbf{P}_i \in \mathbb{R}^{C_s \times C_s}$, where C_s is the length of the atomic property vector and N_i is the number of atoms in the i-th amino acid. By design, amino acid properties are SPD matrices that contain the second-order Riemannian relevance of all atoms regardless of the diversity of input dimensions. These SPD matrices are then processed by SPDNet, enabling Riemannian dimension compression and non-linear mapping. In addition, we record the center of all atomic coordinates of an amino acid as its coordinate, $\mathbf{c}_i \in \mathbb{R}^3$, and combine all amino acid coordinates with their properties as the amino acid sets, $\{(\mathbf{P}_i, \mathbf{c}_i)\}_{i=1}^K$. Next, protein structural hidden features are extracted from the rewrited 3D data through the structure-based extractor.

The structure-based extractor is constructed based on the SPD-GNN block. At the l-th layer of the SPD-GNN block, we dynamically construct the protein as a K-Nearest Neighbors (KNN) graph based on the known amino acid coordinates. $\mathbf{P}_i^l$ is alternately updated as follows:

$$\mathbf{m}_{i,j} = \mathbf{W}_c \begin{bmatrix} \mathbf{P}_i^l & \\ & \mathbf{P}_j^l \end{bmatrix} \mathbf{W}_c^\mathsf{T}, \quad \mathbf{m}_i = \sum_{j \in \mathcal{N}_i} \mathbf{m}_{i,j}, \tag{1}$$

$$\mathbf{P}_i^{l+1} = f_r^l(\mathbf{W}_P \begin{bmatrix} \mathbf{P}_i^l & \\ & \mathbf{m}_i \end{bmatrix} \mathbf{W}_P^\mathsf{T}), \tag{2}$$

where $\mathbf{W}_c \in \mathbb{R}^{C_s \times 2C_s}$ and $\mathbf{W}_P \in \mathbb{R}^{C_s \times 2C_s}$ are the transformation matrices, f_r^l is the ReEig layer, which is introduced by Huang and Van Gool (2017). $\mathbf{m}_{i,j}$ is the association information of the amino acids at both ends of the edge $e_{i,j}$, which is generated based on KNN. We use multiple SPD-GNN layers to obtain the structural hidden features, which are used as pseudo labels to train the sequence-based extractor. Meanwhile, we embed the protein sequence and extract the textual features by DenseNet in the sequence-based extractor. Then, we use SPDNet to compress the covariance of the textual features as required hidden features, $\mathcal{F}_T = \{\mathbf{F}_0, ..., \mathbf{F}_{N_i}\}$, which is aligned with the structural hidden features $\mathcal{F}_S = \{\mathbf{P}_0, ..., \mathbf{P}_{N_i}\}$. Moreover, we design a transformer-based reconstructor to predict the coordinate of each amino acid from the hidden feature, further enhancing its structural information. The extracted structural hidden features are then mapped to Euclidean space, and their dimensions are reduced to vector forms, $\mathbf{F}_p^s \in \mathbb{R}^{L_p \times C_s}$. Last, we concatenate them with the textual features, $\mathbf{F}_p^t$, as the final multi-mode protein features, $\mathbf{F}_p \in \mathbb{R}^{L_p \times (C_t + C_s)}$.

Multi-Mode Drug Feature Extractor

Similar to proteins, we represent drugs in two modalities, textual and structural features. For textual feature extraction, we use the LLM-based Smiles-Bert (Wang et al. 2019) to represent the drug SMILES. Then, the drug textual feature, $\mathbf{F}_d^t \in \mathbb{R}^{L_d \times C_t}$, can be extracted by DenseNet (Hua et al. 2023b) from these representations. For structural feature extraction, we represent the SMILES in the molecular graph, containing the atomic vectors, $\mathbf{F}_d^s \in \mathbb{R}^{N_a \times C_s}$, and the adjacency matrix, $\mathbf{A} \in \mathbb{R}^{N_a \times N_a}$, though Rdkit tools (Landrum et al. 2016). Then, the structural feature is extracted from the molecular graph by our proposed Molecular Graph Attention (M-GAT) block, which is formulated as follows:

$$\mathbf{W}_a = \sigma(\psi(\mathbf{F}_d^{sl})\psi(\mathbf{F}_d^{sl})^\mathsf{T}), \tag{3}$$

$$\mathbf{F}_d^{sl+1} = (\mathbf{W}_a * \mathbf{A})\psi(\mathbf{F}_d^{sl}) + \mathbf{F}_d^{sl}, \tag{4}$$

where $\mathbf{W}_a \in \mathbb{R}^{N_a \times N_a}$ is the atomic relevance matrix, $\sigma()$ and $\psi()$ are the sigmoid and linear layers. Then, we align the structural feature, $\mathbf{F}_d^s \in \mathbb{R}^{L_d \times C_s}$, with the textual feature, and establish an association between them through a cross-attention block, which is formulated as follows:

$$\mathbf{F}_d^{s'} = softmax(\frac{\psi(\mathbf{F}_d^t)\psi(\mathbf{F}_d^s)^\mathsf{T}}{\sqrt{C_s}})\psi(\mathbf{F}_d^s) + \mathbf{F}_d^s. \tag{5}$$

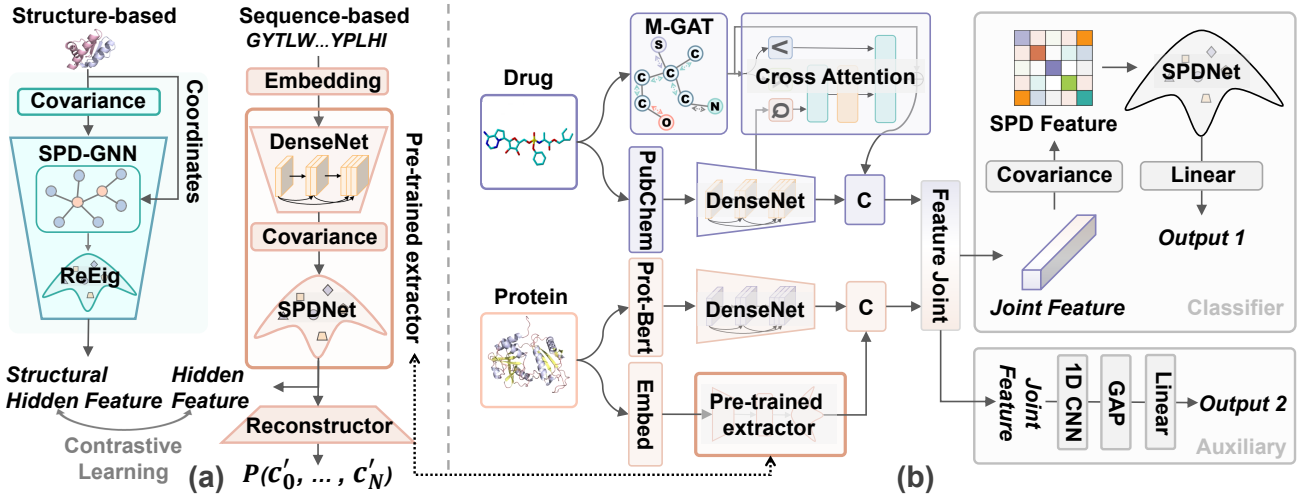

Figure 2: An overview of R-DTI: (a) the pre-trained process of the sequence-based structural feature extractor; (b) the pipeline of drug-target interaction prediction.

Last, we combine the updated structural feature with the textual feature as the final drug features, $\mathbf{F}_d \in \mathbb{R}^{L_d \times (C_t + C_s)}$.

Riemannian Classifier

Before combining the protein and drug features, we propose a new method to unify their feature space. As shown in Figure 1 (f), we embed the protein and drug features in a new feature space, $\mathbb{V}$, where they obey strict orthogonality. $\mathbb{V}_P \perp \mathbb{V}_D$. Then, we concatenate the protein and drug features and map them into a unified space, the process could be formulated as follows:

$$\mathbf{F}_{joint} = \psi(\begin{bmatrix} \mathbf{F}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_d \end{bmatrix}), \tag{6}$$

where $\mathbf{F}_{joint} \in \mathbb{R}^{L_j \times C_j}$ is the joint feature, L_j and C_j are its length and channel, respectively. $\psi()$ is the mapping process though linear and GELU layers. Then, we compute the covariance matrix of the joint feature in its two dimensions as its Riemannian features, the channel-based $\mathbf{F}_{r^C} \in \mathbb{R}^{C_j \times C_j}$ and the length-based $\mathbf{F}_{r^L} \in \mathbb{R}^{L_j \times L_j}$.

To compress these Riemannian features into a stable feature space, we use the SPDNet, which consists of BiMap, ReEig, and LogEig layers, to enhance the most relevant information for DTI prediction. The three layers of SPDNet are introduced in ***Appendix A***. The elements of the upper triangle of the compressed feature are transformed into vectors as the interactive features of the protein-drug pairs. Finally, we predict the DTI from these interactive features by the Multi-Layer Perceptron (MLP) layers. Note that while the Riemannian feature contains the second-order relevance of the sample pairs, which is beneficial for DTI prediction, the relevance calculation complicates the gradient computation and hinders the training of feature extractors. Therefore, we proposed an auxiliary classifier composed of 1D convolution and MLP layers to directly backwards the semantic information to the above two multi-mode feature extractors, as shown in Figure 2. By design, the extracted features are highly correlated with the task, which further improves the performance of R-DTI. Our loss functions, hyperparameter settings and measure metrics are shown in ***Appendix B***.

Evaluation

Evaluation Datasets

We evaluate the proposed method on six datasets, including four classical DTI benchmarks, Human, C.elegans (Tsubaki, Tomii, and Sese 2019), BindingDB (Chen et al. 2020), DrugBank (Zhao et al. 2022), and two DTA datasets, Davi and KIBA (Davis et al. 2011). The statistics of the six datasets are listed in ***Appendix C***, and their basic information is reported in Figure 3. Specifically, Figure 3 (a) presents the common issue that the length of the protein is much longer than the drug molecule. Figure 3 (b) shows the number of positive and negative samples for the six datasets. We visualize the distribution of affinity labels for two DTA datasets in Figure 3 (c). To evaluate the accuracy of the DTI prediction models during label imbalance, we reconstruct the two DTA datasets as binary classification datasets. Consistent with (Zhao et al. 2022), we set the thresholds to 5.0 and 12.1 for the Davis and KIBA datasets, respectively. Besides, to evaluate the generalization of DTI models, we split the test data of BindingDB (Chen et al. 2020) into three subsets, ***Unseen drug***, ***Unseen target***, and ***Unseen pair***. The number of positive and negative samples is shown in Figure 3 (d).

Comparison with the State-of-The-Art

We compare the proposed R-DTI method with the state-of-the-art DTI prediction models, including TransformerCPI (Chen et al. 2020), HADTI (Zhao et al. 2022), CPInformer (Hua et al. 2023b), TriMulDTI (Dehghan et al. 2023), MFR-DTA (Hua et al. 2023a), and PSC-CPI (Wu et al. 2024). We report the experimental results (measured by AUC-ROC and AUC-PR) for four classical DTI benchmarks in Table 1. The proposed R-DTI method performs better than all other methods in most metrics, especially in the BindingDB and Drug-

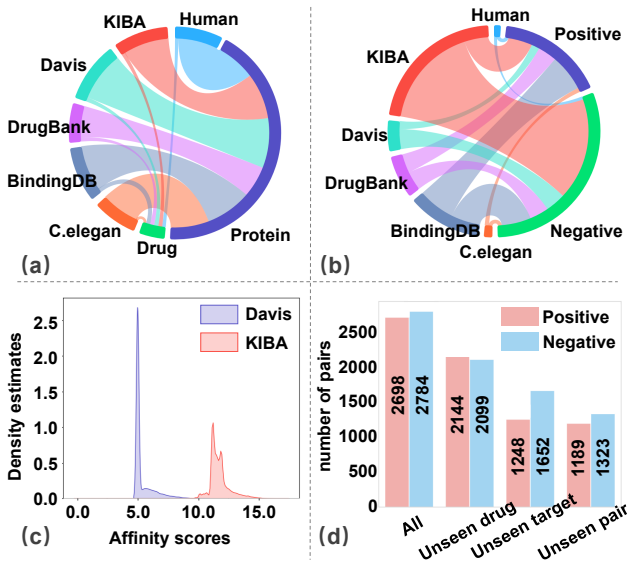

Figure 3: Dataset information: (a) the average length of proteins and drugs; (b) the number of positive and negative samples; (c) the distribution of affinity scores in Davis and KIBA; (d) the numbers of positive and negative samples of three subsets in BindingDB.

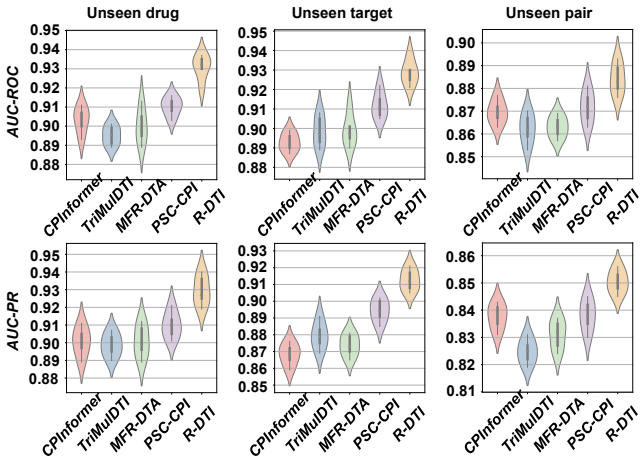

Figure 4: A comparison with the state-of-the-art methods for the unseen drug, target, and pair settings in BindingDB.

Bank datasets. From Figure 3 (b), we can see that the quality of BindingDB and DrugBank is higher, with a larger amount of data and balanced labels, which is more suitable for evaluating the overall performance of DTI models. Specifically, the proposed R-DTI method can achieve 1.4% and 1.0% performance gains in terms of AUC-ROC and AUC-PR, respectively, as compared to the second best method, PSC-CPI, in the BindingDB. Meanwhile, in the DrugBank, the AUC-ROC and AUC-PR of our method are 1.9% and 1.5% higher than PSC-CPI, respectively, demonstrating the superiority of R-DTI. In small-sample datasets, Human and C.elegans, the superior performance of the R-DTI model remains stable.

To evaluate the generalization, in Figure 4, we present the performance of the state-of-the-art methods on the three test subsets of BindingDB, including *Unseen drug*, *Unseen target*, and *Unseen pair*. We find that most of these methods perform worse in the *Unseen target* subset than in the *Unseen drug* subset, indicating that their learned parameters are biased towards the proteins to some extent. However, R-DTI using Riemannian SPD features, which is the covariance matrix of the joint protein and drug features, can effectively improve its generalization in the *Unseen target* subset. In addition, the *Unseen pair* subset is the most difficult setting, but the R-DTI model has a remarkable performance in terms of both AUC-ROC and AUC-PR, with a performance gain of 1.2% compared to the second best method, PSC-CPI.

For two DTA datasets, Davis and KIBA, we evaluate existing methods by two tasks, interaction (measured by AUC-ROC and AUC-PR) and affinity prediction (measured by CI and MSE), and report the results in Table 2. The label distribution of both datasets is unbalanced, so they can be effectively used to evaluate the robustness of the model in dealing with the label imbalance classification and the long-tailed regression tasks. Table 2 reports that all methods fail to achieve satisfactory performance in terms of AUC-PR on both datasets, which also confirms the above viewpoint. However, the performance of the R-DTI model is comparatively robust in these metrics, further demonstrating its merits. In addition, our proposed method also performs well in affinity prediction tasks, and it achieves the best results in terms of both CI and MSE on both two DTA datasets, verifying its robustness for the long-tailed regression task.

Ablation Studies and Discussion

Component Analysis The R-DTI model is our newly developed framework without relying on any existing baseline method. Therefore, we construct a basic DTI prediction architecture as a baseline (Model-1) to evaluate the innovations on two classical DTI benchmarks, BindingDB and DrugBank. Specifically, the baseline consists of the textural protein feature extractor and the complete multi-mode drug feature extractor, and applies concatenation to joint protein and drug features. On this basis, the auxiliary classifier is used to predict the DTI. We report the results of the models with different configurations of our novel components in Table 3. It is clear that Model-2, which adds protein structural features, performs better than Model-1 on both datasets, indicating the effectiveness of the proposed extractor. Model-3 and Model-4 show the results of using only the feature joint block or the Riemann classifier based on Model-2, respectively, and both show some performance improvement, confirming the benefits of both components. However, we note that the combination of the two components, *i.e.*, Model-5, shows a huge improvement in accuracy on both datasets. This improvement largely confirms that the performance of the Riemannian classifier is limited when the feature spaces of proteins and drugs are not uniform, and the excellent ability of the classifier depends on the feature joint block. We also test the Riemannian classifier using only general drug and protein features as Model-6, and the result shows that

Method	BindingDB		DrugBank		Human		C.elegans	
	***AUC-ROC*↑**	***AUC-PR*↑**	***AUC-ROC*↑**	***AUC-PR*↑**	***AUC-ROC*↑**	***AUC-PR*↑**	***AUC-ROC*↑**	***AUC-PR*↑**
TransformerCPI	$0.951_{\pm 0.003}$	$0.949_{\pm 0.004}$	$0.837_{\pm 0.003}$	$0.836_{\pm 0.002}$	$0.973_{\pm 0.002}$	$0.979_{\pm 0.002}$	$0.988_{\pm 0.002}$	$0.987_{\pm 0.001}$
HADTI	$0.956_{\pm 0.003}$	$0.951_{\pm 0.004}$	$0.889_{\pm 0.001}$	$0.897_{\pm 0.001}$	$0.978_{\pm 0.002}$	$0.980_{\pm 0.003}$	$0.985_{\pm 0.003}$	$0.987_{\pm 0.003}$
CPInformer	$0.965_{\pm 0.002}$	$0.966_{\pm 0.003}$	$0.887_{\pm 0.002}$	$0.891_{\pm 0.002}$	$0.985_{\pm 0.003}$	$0.989_{\pm 0.002}$	$0.991_{\pm 0.002}$	$0.992_{\pm 0.001}$
TriMulDTI	$0.963_{\pm 0.003}$	$0.967_{\pm 0.002}$	$0.895_{\pm 0.002}$	$0.897_{\pm 0.003}$	$0.985_{\pm 0.003}$	$0.986_{\pm 0.002}$	$0.986_{\pm 0.003}$	$0.988_{\pm 0.002}$
MFR-DTA	$0.962_{\pm 0.002}$	$0.959_{\pm 0.003}$	$0.893_{\pm 0.002}$	$0.891_{\pm 0.002}$	$0.986_{\pm 0.003}$	$0.983_{\pm 0.002}$	$0.988_{\pm 0.002}$	$0.985_{\pm 0.001}$
PSC-CPI	$0.969_{\pm 0.001}$	$0.972_{\pm 0.002}$	$0.895_{\pm 0.001}$	$0.897_{\pm 0.001}$	$0.988_{\pm 0.002}$	$\mathbf{0.993}_{\pm 0.002}$	$0.993_{\pm 0.001}$	$0.992_{\pm 0.001}$
R-DTI	$\mathbf{0.983}_{\pm 0.001}$	$\mathbf{0.982}_{\pm 0.002}$	$\mathbf{0.914}_{\pm 0.001}$	$\mathbf{0.912}_{\pm 0.001}$	$\mathbf{0.991}_{\pm 0.001}$	$0.992_{\pm 0.003}$	$\mathbf{0.996}_{\pm 0.001}$	$\mathbf{0.996}_{\pm 0.001}$

Table 1: A comparison with the SOTA methods in terms of AUC-ROC and AUC-PR on the four classical DTI benchmarks.

Method	$\text{Davis}_{\text{(interaction)}}$		$\text{Davis}_{\text{(affinity)}}$		$\text{KIBA}_{\text{(interaction)}}$		$\text{KIBA}_{\text{(affinity)}}$	
	***AUC-ROC*↑**	***AUC-PR*↑**	***CI*↑**	***MSE*↓**	***AUC-ROC*↑**	***AUC-PR*↑**	***CI*↑**	***MSE*↓**
TransformerCPI	$0.878_{\pm 0.001}$	$0.767_{\pm 0.002}$	$0.859_{\pm 0.003}$	$0.286_{\pm 0.002}$	$0.900_{\pm 0.001}$	$0.785_{\pm 0.002}$	$0.859_{\pm 0.002}$	$0.203_{\pm 0.002}$
HADTI	$0.920_{\pm 0.002}$	$0.838_{\pm 0.002}$	$0.876_{\pm 0.001}$	$0.279_{\pm 0.001}$	$0.916_{\pm 0.001}$	$0.799_{\pm 0.002}$	$0.871_{\pm 0.003}$	$0.191_{\pm 0.003}$
CPInformer	$0.921_{\pm 0.001}$	$0.833_{\pm 0.003}$	$0.874_{\pm 0.002}$	$0.277_{\pm 0.002}$	$0.923_{\pm 0.002}$	$0.814_{\pm 0.001}$	$0.867_{\pm 0.003}$	$0.183_{\pm 0.001}$
TriMulDTI	$0.924_{\pm 0.002}$	$0.847_{\pm 0.001}$	$0.896_{\pm 0.002}$	$0.229_{\pm 0.003}$	$0.927_{\pm 0.002}$	$0.814_{\pm 0.003}$	$0.893_{\pm 0.003}$	$0.145_{\pm 0.002}$
MFR-DTA	$0.926_{\pm 0.003}$	$0.845_{\pm 0.002}$	$0.905_{\pm 0.001}$	$0.221_{\pm 0.001}$	$0.928_{\pm 0.001}$	$0.816_{\pm 0.002}$	$0.898_{\pm 0.002}$	$0.136_{\pm 0.001}$
PSC-CPI	$0.929_{\pm 0.002}$	$0.850_{\pm 0.002}$	$0.903_{\pm 0.001}$	$0.217_{\pm 0.001}$	$\mathbf{0.936}_{\pm 0.001}$	$0.824_{\pm 0.001}$	$0.901_{\pm 0.001}$	$0.139_{\pm 0.001}$
R-DTI	$\mathbf{0.937}_{\pm 0.001}$	$\mathbf{0.868}_{\pm 0.002}$	$\mathbf{0.911}_{\pm 0.001}$	$\mathbf{0.212}_{\pm 0.001}$	$0.935_{\pm 0.001}$	$\mathbf{0.830}_{\pm 0.002}$	$\mathbf{0.906}_{\pm 0.001}$	$\mathbf{0.127}_{\pm 0.001}$

Table 2: A comparison with the state-of-the-art in terms of AUC-ROC, AUC-PR, CI, and MSE on the **Davis** and **KIBA**.

Model	Protein structural feature extractor	Feature joint block	Riemannian classifier	Auxiliary classifier	BindingDB		DrugBank	
					***AUC-ROC*↑**	***AUC-PR*↑**	***AUC-ROC*↑**	***AUC-PR*↑**
Model-1	✗	✗	✗	✓	$0.952_{\pm 0.003}$	$0.950_{\pm 0.004}$	$0.857_{\pm 0.004}$	$0.855_{\pm 0.003}$
Model-2	✓	✗	✗	✓	$0.960_{\pm 0.002}$	$0.961_{\pm 0.002}$	$0.871_{\pm 0.003}$	$0.873_{\pm 0.003}$
Model-3	✓	✓	✗	✓	$0.965_{\pm 0.001}$	$0.968_{\pm 0.002}$	$0.883_{\pm 0.001}$	$0.886_{\pm 0.001}$
Model-4	✓	✗	✓	✗	$0.963_{\pm 0.001}$	$0.967_{\pm 0.002}$	$0.884_{\pm 0.001}$	$0.885_{\pm 0.001}$
Model-5	✓	✓	✓	✗	$0.977_{\pm 0.001}$	$0.979_{\pm 0.002}$	$0.906_{\pm 0.001}$	$0.908_{\pm 0.001}$
Model-6	✗	✓	✓	✗	$0.962_{\pm 0.002}$	$0.963_{\pm 0.002}$	$0.891_{\pm 0.001}$	$0.894_{\pm 0.001}$
R-DTI	✓	✓	✓	✓	$\mathbf{0.983}_{\pm 0.001}$	$\mathbf{0.982}_{\pm 0.002}$	$\mathbf{0.914}_{\pm 0.001}$	$\mathbf{0.912}_{\pm 0.001}$

Table 3: Ablation study results in terms of AUC-ROC and AUC-PR on the **BindingDB** and **DrugBank**.

it is suitable for general situation. Despite the better predictive performance of the Riemannian classifiers, the complicated gradient computation from the second-order relevance calculation affects the parameter optimization of their upstream feature extractors. Therefore, we introduced an auxiliary classifier to help train the feature extractors, and the final model obtained, *i.e.*, R-DTI, achieves the best performance.

Analysis of Protein Structural Feature Extraction To prove the effectiveness of Riemannian networks in extracting hidden structural features of proteins, we compare the distribution of the extracted features in Figure 5 (a). We first cluster the protein samples, which are not involved during training, into four classes based on the 3D structural representation, $\mathcal{P}^a = \{(\mathbf{p}_i^a, \mathbf{c}_i^a)\}_{i=1}^N$. These four colors indicate different proteins to analyze the distribution of the corresponding features. Obviously, the sequence representations are disordered due to the lack of structural information in the initial sequence embedding. To complement this information, we train the sequence-based feature extractor through contrastive and reconstructive learning, and form the hidden features into two spaces, Euclidean and Riemannian SPD.

The Riemannian SPD features are extracted by the pre-trained extractor mentioned above, and the Euclidean features are extracted by a similar extractor, that uses zero padding to make the amino acid structure representations equal in length. Figure 5 (a) shows that both methods can assist the sequence-based extractor in learning structural information. However, the Euclidean hidden feature scatter is slightly distorted by the original structural features, *i.e.*, there is a bias in the structural information when the protein features are transformed by the Euclidean extractor. In contrast, the Riemannian SPD hidden features are more consistent to the distribution of the structural features. The proposed pre-trained extractor uses the second-order relevance of the atom vector to represent the amino acid, thus compressing the representation into a more stable Riemannian space. Moreover, we use the ReEig layer f_r^l to ensure

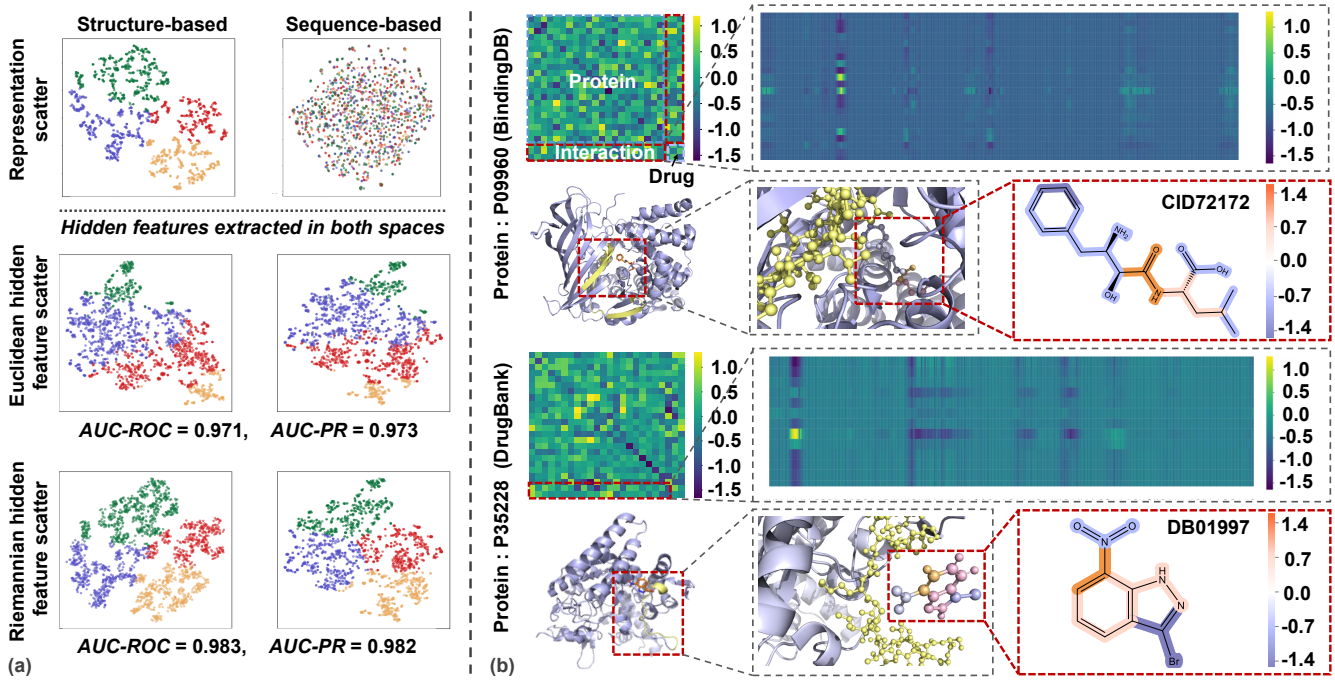

Figure 5: Left (a): Visualization comparison of protein hidden features extracted in Euclidean and Riemannian space. Right (b): Visualization of the joint SPD features and the response between protein amino acids and drug atoms.

that the transformed features are symmetric positive definite, which prevents feature space shifts due to changes in their rank. Therefore, the Riemannian SPD hidden features contribute to better performance on BindingDB, with a 1.2% and 0.9% improvement in AUC-ROC and AUC-PR.

Analysis of the Joint Features and Visualization of DTI From the joint feature, we extract two second-order Riemannian relevancies, including the channel-based $\mathbf{F}_{r^C}$ and the length-based $\mathbf{F}_{r^L}$ Riemannian features. The role of channel-based features is intuitive, since each of them is determined by the joint action of protein and drug features. Therefore, during the inference process, the parameters of the decision model are not biased toward either the protein or the drug, which is effective in increasing the robustness of the model. In parallel, the length-based feature can further enhance the second-order relevance between protein and drug, while improving the interpretability of the decision.

Based on the length-based Riemannian feature, we visualize two pairs of samples from BindingDB and DrugBank, in Figure 5 (b). It is easy to observe the interactive degree between the amino acids in the protein and the atoms in the drug by visualizing the length-based Riemannian feature, $\mathbf{F}_{r^L} \in \mathbb{R}^{L_j \times L_j}$, and its compressed feature, $\mathbf{F}'_{r^L} \in \mathbb{R}^{C_{k^L} \times C_{k^L}}$. The compressed feature heatmap can provide a global sense of the DTI, including the covariance matrices of both the protein and the drug itself and their interaction, and the length-based feature can magnify the interaction part to focus on the interactive relationship between the protein and the drug. On this basis, we visualize the structure of the protein and the drug, highlighting their high response regions in yellow and orange, respectively. Although these high-response regions are not related to the actual binding sites or functional groups, they could also help us find the region of interest from the deep model. Interestingly, the covariance matrix allows the model to learn the region of interest more directly than the attention mechanism.

Conclusion

We present a novel R-DTI model based on second-order relevance exploration for better feature extraction and prediction formulation. First, we construct a sequence-based protein structural feature extractor that can extract spatial-preserved structural features from pre-trained prior knowledge. Notably, the extracted feature contains the second-order relevance of atoms, which represents protein structural information more comprehensively and stably. Then, we develop a Riemannian classifier that avoids the parameter-biased protein problem while clarifying the semantic information of DTI features, thus making it more stable. We also visualize the second-order Riemannian relevance and the high-response regions of biological molecules, to explore the correlation between them. The experimental results confirm the superiority of R-DTI over the state-of-the-art approaches in six benchmark datasets. However, there is still room for improvement, such as designing a better Riemannian representation of proteins and more efficient DTI models. In the future, we will further explore the second-order relevance in feature extraction and prediction formulation of DTI prediction to empower AI for drug discovery.

Acknowledgments

This work was supported by the National Key R&D Program of China (2023YFE0116300, 2023YFF1105102, 2023YFF1105105), the National Natural Science Foundation of China (Grant NO. 62106089, 62336004), the Major Project of the National Social Science Foundation of China (No. 21&ZD166) and the Natural Science Foundation of Jiangsu Province (No. BK20221535).

References

Abbasi, K.; Razzaghi, P.; Poso, A.; Amanlou, M.; Ghasemi, J. B.; and Masoudi-Nejad, A. 2020. DeepCDA: deep cross-domain compound–protein affinity prediction through LSTM and convolutional neural networks. *Bioinformatics*, 36(17): 4633–4642.

Arsigny, V.; Fillard, P.; Pennec, X.; and Ayache, N. 2007. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM journal on matrix analysis and applications*, 29(1): 328–347.

Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; and Bourne, P. E. 2000. The protein data bank. *Nucleic acids research*, 28(1): 235–242.

Cao, D.-S.; Liu, S.; Xu, Q.-S.; Lu, H.-M.; Huang, J.-H.; Hu, Q.-N.; and Liang, Y.-Z. 2012. Large-scale prediction of drug–target interactions using protein sequences and drug topological structures. *Analytica chimica acta*, 752: 1–10.

Chen, L.; Tan, X.; Wang, D.; Zhong, F.; Liu, X.; Yang, T.; Luo, X.; Chen, K.; Jiang, H.; and Zheng, M. 2020. TransformerCPI: improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. *Bioinformatics*, 36(16): 4406–4414.

Coelho, E. D.; Arrais, J. P.; and Oliveira, J. L. 2016. Computational discovery of putative leads for drug repositioning through drug-target interaction prediction. *PLoS computational biology*, 12(11): e1005219.

Davis, M. I.; Hunt, J. P.; Herrgard, S.; Ciceri, P.; Wodicka, L. M.; Pallares, G.; Hocker, M.; Treiber, D. K.; and Zarrinkar, P. P. 2011. Comprehensive analysis of kinase inhibitor selectivity. *Nature biotechnology*, 29(11): 1046–1051.

Dehghan, A.; Razzaghi, P.; Abbasi, K.; and Gharaghani, S. 2023. TripletMultiDTI: Multimodal Representation Learning in Drug-Target Interaction Prediction with Triplet Loss Function. *Expert Systems with Applications*, 120754.

Detlefsen, N. S.; Hauberg, S.; and Boomsma, W. 2022. Learning meaningful representations of protein sequences. *Nature communications*, 13(1): 1914.

Elnaggar, A.; Heinzinger, M.; Dallago, C.; Rehawi, G.; Wang, Y.; Jones, L.; Gibbs, T.; Feher, T.; Angerer, C.; Steinegger, M.; et al. 2021. Prottrans: Toward understanding the language of life through self-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 44(10): 7112–7127.

Helgason, S. 1965. Radon-Fourier transforms on symmetric spaces and related group representations. *Bulletin of the American Mathematical Society*, 71(5): 757–763.

Hua, Y.; Feng, Z.; Song, X.; Li, H.; Xu, T.; Wu, X.-J.; and Yu, D.-J. 2024. APMG: 3D Molecule Generation Driven by Atomic Chemical Properties. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

Hua, Y.; Feng, Z.; Song, X.; Wu, X.-J.; and Kittler, J. 2025. MMDG-DTI: Drug–target interaction prediction via multimodal feature fusion and domain generalization. *Pattern Recognition*, 157: 110887.

Hua, Y.; Song, X.; Feng, Z.; and Wu, X. 2023a. MFR-DTA: a multi-functional and robust model for predicting drug–target binding affinity and region. *Bioinformatics*, 39(2): btad056.

Hua, Y.; Song, X.; Feng, Z.; Wu, X.-J.; Kittler, J.; and Yu, D.-J. 2023b. CPInformer for Efficient and Robust Compound-Protein Interaction Prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 20(1): 285–296.

Huang, Z.; and Van Gool, L. 2017. A riemannian network for spd matrix learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31.

Jacob, L.; and Vert, J.-P. 2008. Protein-ligand interaction prediction: an improved chemogenomics approach. *Bioinformatics*, 24(19): 2149–2156.

Landrum, G.; et al. 2016. Rdkit: Open-source cheminformatics software. 2016. *doi http://www. rdkit. org/, https://github. com/rdkit/rdkit*, 149(150): 650.

Lee, I.; Keum, J.; and Nam, H. 2019. DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences. *PLOS Computational Biology*, 15(6): 1–21.

Lim, J.; Ryu, S.; Park, K.; Choe, Y. J.; Ham, J.; and Kim, W. Y. 2019. Predicting drug–target interaction using a novel graph neural network with 3D structure-embedded graph representation. *Journal of chemical information and modeling*, 59(9): 3981–3988.

Öztürk, H.; Özgür, A.; and Ozkirimli, E. 2018. DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics*, 34(17): i821–i829.

Song, Y.; Wang, X.; Yao, J.; Liu, W.; Zhang, J.; and Xu, X. 2024. ViTGaze: Gaze Following with Interaction Features in Vision Transformers. *Visual Intelligence*, 1(2): 31.

Tang, J.; Szwajda, A.; Shakyawar, S.; Xu, T.; Hintsanen, P.; Wennerberg, K.; and Aittokallio, T. 2014. Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. *Journal of Chemical Information and Modeling*, 54(3): 735–743.

Tang, J.; Wang, J.; and Hu, J.-F. 2023. Predicting human poses via recurrent attention network. *Visual Intelligence*, 1(1): 18.

Tsubaki, M.; Tomii, K.; and Sese, J. 2019. Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics*, 35(2): 309–318.

Wang, S.; Guo, Y.; Wang, Y.; Sun, H.; and Huang, J. 2019. Smiles-bert: large scale unsupervised pre-training for molecular property prediction. In *Proceedings of the 10th ACM international conference on bioinformatics, computational biology and health informatics*, 429–436.

Wang, Y.; Shen, Y.; Chen, S.; Wang, L.; Fei, Y.; and Zhou, H. 2022. Learning Harmonic Molecular Representations on Riemannian Manifold. In *The Eleventh International Conference on Learning Representations*.

Wu, L.; Huang, Y.; Tan, C.; Gao, Z.; Hu, B.; Lin, H.; Liu, Z.; and Li, S. Z. 2024. PSC-CPI: Multi-Scale Protein Sequence-Structure Contrasting for Efficient and Generalizable Compound-Protein Interaction Prediction. In *AAAI Conference on Artificial Intelligence*.

Xu, T.; Zhu, X.-F.; and Wu, X.-J. 2023. Learning spatio-temporal discriminative model for affine subspace based visual object tracking. *Visual Intelligence*, 1(1): 4.

Yamanishi, Y.; Araki, M.; Gutteridge, A.; Honda, W.; and Kanehisa, M. 2008. Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*, 24(13): i232–i240.

Zhao, Q.; Zhao, H.; Zheng, K.; and Wang, J. 2022. HyperAttentionDTI: improving drug–protein interaction prediction by sequence-based deep learning with attention mechanism. *Bioinformatics*, 38(3): 655–662.