Pattern Recognition 163 (2025) 111463

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

MDFCL: Multimodal data fusion-based graph contrastive learning framework for molecular property prediction

Xu Gong [a], Maotao Liu [a], Qun Liu [a,*], Yike Guo [b], Guoyin Wang [a]

[a] Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Nan'an District, 400065, Chongqing, China
[b] The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, 999077, Hong Kong, China

ARTICLE INFO

Keywords:
Molecular property prediction
Graph representation
Multimodal data fusion
Graph contrastive learning

ABSTRACT

Molecular property prediction is a critical task with substantial applications for drug design and repositioning. The multiplicity of molecular data modalities and paucity of labeled data present significant challenges that affect algorithmic performance in this domain. Nevertheless, conventional approaches typically focus on singular data modalities and ignore either hierarchical structural features or other data pattern information, leading to problems when expressing complex phenomena and relationships. Additionally, the scarcity of labeled data obstructs the accurate mapping of instances to labels in property prediction tasks. To address these issues, we propose the **M**ultimodal **D**ata **F**usion-based graph **C**ontrastive **L**earning framework (MDFCL) for molecular property prediction. Specifically, we incorporate exhaustive information from dual molecular data modalities, namely graph and sequence structures. Subsequently, adaptive data augmentation strategies are designed based on the molecular backbones and side chains for multimodal data. Built upon these augmentation strategies, we develop a graph contrastive learning framework and pre-train it with unlabeled data (~ 10M molecules). MDFCL is tested using 13 molecular property prediction benchmark datasets, demonstrating its effectiveness through empirical findings. In addition, a visualization study demonstrates that MDFCL can embed molecules into representative features and steer the distribution of molecular representations.

1. Introduction

Increasingly, deep learning has become a powerful technique for analyzing and processing molecule data in bioinformatics and chem-informatics [1]. Various sophisticated intelligent models have made significant progress in molecular property prediction, docking simulation, retrosynthesis tasks, and similar applications. However, the potential of these models is limited by the scale of labeled data, and it is time-consuming and laborious to collect the labels of molecules, with the process requiring tedious wet experiments and expert knowledge for validation. In particular, the scarcity of labeled data poses a challenge in accurately establishing the mapping of instances to labels in molecular property prediction. Building upon the significant success of pre-training strategies in natural language processing (NLP) and computer vision [2], numerous chemical pre-training models have emerged to learn molecular representations by mining inherent chemical semantics with unlabeled data [3].

Molecules contain multiple modal forms, whereas graphs serve as intuitive mathematical topology representations. Graph neural networks (GNN) can efficiently extract features from molecular graph data. The early development of GNN built on spectral-based formulation, which utilizes spectral graph theory to generalize convolutional operations to graphs, termed spectral convolution methods [4,5]. To achieve scalability for large-scale graphs, spatial approaches [5,6] directly aggregate information from neighboring nodes. Attention-based variants, such as GAT [7] and GANM [8], utilize attention weights to effectively capture the topological structures and dependencies between nodes. It is crucial for perceiving functional group structures contained in molecules. In addition, gating mechanism-based and hierarchical GNN [9] have been developed. These GNN architectures have made significant success in handling molecular graph modality [10–12]. The simplified molecular input line entry system (SMILES) encodes molecular structures into string sequences using predefined rules. In contrast, the accurate measurement of 3D information and molecular conformation is expensive and impractical.

Hence, with a primary focus on the sequence and graph modalities, there are two main research directions for molecular pre-training. For sequence-based pre-training, Molformer [13], KPGT [14], and KV-PLM [15] randomly mask several tokens in the SMILES strings and input the broken SMILES strings into Transformer to reconstruct them.

* Corresponding author.
E-mail address: liuqun@cqupt.edu.cn (Q. Liu).

https://doi.org/10.1016/j.patcog.2025.111463
Received 29 September 2023; Received in revised form 4 February 2025; Accepted 9 February 2025
Available online 16 February 2025

Table 1
Comparison between related works and the proposed MDFCL.

Method	Sequence	Graph	Adaptability	Objective	Model
Molformer [13]	✓	✗	✗	MCM	Transformer
KPGT [14]	✓	✗	✗	MCM	LiGhT
HierMRL [17]	✗	✓	✓	SSC	GNN
MolCLR [11]	✗	✓	✗	SSC	GNN
Ours	✓	✓	✓	SSC	GNN+CNN

The pre-training objective of these methods is referred to as masked component modeling (MCM) [3]. However, one study [10] argues that the masking mechanism is unreasonable due to the imbalance and limited atoms set in nature. For graph-based pre-training, similar to the MCM objective in sequence modality, prior studies [16] have employed universal GNN frameworks as skeletons to predict masked nodes and substructures. In contrast, some studies [17] focus on overall molecular structures to accomplish the pre-training by performing the same-scale contrast (SSC) of individual molecular graphs. The objective is to design various augmentation strategies that generate augmentation instances and then push the positive instances close to the anchor molecule and away from the negative instances, as in the case of MolCLR [11] and its various variants [18,19].

Although these single-modality pre-training methods effectively alleviate the issue of limited label data, they inherently possess certain deficiencies. SMILES sequences are adept at encapsulating the linear sequences of molecules; however, they struggle to preserve the ring structures and tree-like branching of graph data. It is also difficult to thoroughly represent molecular features solely through single-molecular data modality. The fusion of modal information can offer substantial complementary advantages to significantly enhance molecular representation [14]. Therefore, the integration of sequence and graph modalities of molecules is critical.

To tackle the issues above, inspired by MolCLR [11] and works [2, 17], we propose a multimodal data fusion-based graph contrastive learning framework (MDFCL), as shown in Fig. 1. The closest to our work is MolCLR [11], which learns representations through contrastive learning of molecular graphs and catalyzes a line of subsequent study [3]. Considering the unique characteristics of molecules, our study further develops adaptive augmentation strategies and integrates multimodal data fusion in this context. A comparison between MDFCL and related methods is presented in Table 1. All code is released at repository[1]. On the whole, our contributions are as follows:

- We propose a novel self-supervised framework for molecular property prediction, dubbed MDFCL, which effectively integrates SMILE sequences and molecular graphs to explore more informative and discriminative representations, alleviating the limitations of label scarcity and insufficient information in single-modal data.
- Based on the backbone and side chains, we develop adaptive augmentation strategies to empower MDFCL to concentrate more intensely on critical molecular structural information. The strategies take a backbone as the core to generate valid augmentation instances by reorganizing side chains, facilitating the processing capability and prediction precision of complex chemical reactions.
- Built upon the above strategies, we extend and optimize the conventional graph contrastive learning paradigm to pre-train for multi-modal unlabeled molecular data. It constructs four kinds of augmentation instances and three hierarchical losses to promote learning a refined and precise chemical space.
- Extensive experiments are conducted on 13 benchmark datasets encompassing 61 binary classification tasks and 24 regression tasks. The results demonstrate that the proposed MDFCL achieves competitive performance.

[1] https://github.com/lukcats/MDFCL

2. Related works

2.1. Molecular representation leaning

Various deep models trained on molecular graphs generally achieve high performance in molecular representation learning. Some generic GNN frameworks have been redesigned and applied directly to universal molecular graphs, such as HierMRL [17] and MolCLR [11], in which bonds and atoms are represented by edges and nodes, respectively. Other researchers have focused on constructing more informative molecular graphs, such as motif_graph [20], N_Gram graph [18], and non-polar covalent bond graph [21]. These graphs are also fed into GNNs to yield promising performance. To capture functional group information, the substructure attention [22] and hierarchical learning mechanism [23,24] have been explored to enhance molecular representation. Hierarchical learning has also contributed to the chemical semantic interpretation of model predictions in [25]. Although these models have achieved competitive performance, the issue of limited labeled data remains a challenge.

2.2. Molecular pre-training

Molecular pre-training leverages unlabeled data, enabling the model to learn latent patterns and intrinsic semantics of molecules. Inspired by NLP, some researchers treat molecular SMILES as natural languages and apply NLP pre-training techniques such as SMILES-BERT [26], Molformer [13], and Chemformer [27] to process them. Under this approach, the pre-training objective is to recover masked elements in SMILES sequences using contextual information. Additional pre-training objectives, such as the chemical fingerprint alignment of molecules [28] or correlation prediction of chemical bond [29], can also be defined. MolCLR [11] and its variants [17,23] extend the molecular pre-training to consider interactions between individual molecular graphs. These graph-based pre-training models adopt contrast learning as an objective that maximizes anchor sample similarity to positive samples while minimizing similarity to negative samples. This extension represents a significant advancement, as it prioritizes the overall molecular structure and graph modality. Additional types of information used in pre-training, such as textual and conformation information, as employed by [15,30], are beyond the scope of our work. Learning molecular features solely through a single data modality poses challenges for these pre-training methods.

3. Method

3.1. Notations and problem definition

In MDFCL, each molecule contains inputs representing graph modality G and sequence modality S. We define G as (V, E), where $V = \{v_1, v_2, \ldots, v_n\}$ denotes the set of nodes depicted by a feature matrix $X \in \mathbb{R}^{n\times d}$ and E denotes the set of edges depicted by an adjacent matrix $A \in \mathbb{R}^{n\times n}$. We also define S as $\{s_1, s_2, \ldots, s_n\}$, where s_i represents the ith character in the SMILES string and n denotes the length of the SMILES string. G and S are encoded by the graph encoder φ_g and sequence encoder φ_s, respectively. φ_g and φ_s constitute the molecular multimodal encoder φ that generates the feature representations of molecules. Our objective is to learn the salient features and patterns among molecules by pre-training φ on the unlabeled molecule set $\mathcal{M}_u$. We then fine-tune the model using the labeled sample set $\mathcal{M}_l$ for property prediction. The problem above can be formulated as follows:

$$\varphi_{\text{pre}} = \text{Pretrain}\left(\varphi, \mathcal{M}_u^1, \mathcal{M}_u^2\right), \tag{1}$$

$$\varphi_{\text{fine}} = \text{Fine-tune}\left(\varphi_{\text{pre}}, \mathcal{M}_l^1, \mathcal{M}_l^2\right), \tag{2}$$

where $\mathcal{M}_l^1$ and $\mathcal{M}_l^2$ denote two modalities of the labeled data.

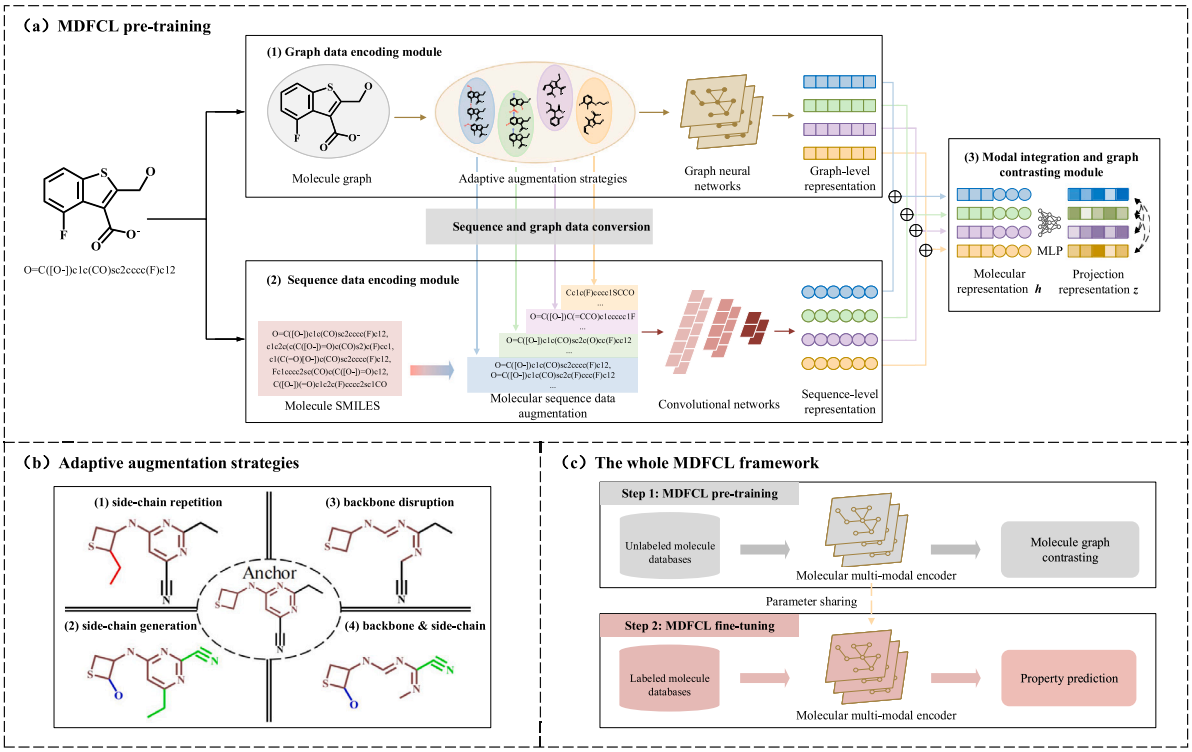

Fig. 1. Overview of MDFCL. (a) MDFCL pre-training. (b) Molecular adaptive augmentation strategies: side-chain repetition, backbone disruption, and side-chain generation. (c) The whole MDFCL framework: the molecular multimodal encoder is first pre-trained and then fine-tuned for downstream property prediction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. Overview of MDFCL

As visualized in Fig. 1, MDFCL is a novel semi-supervised framework trained on large multimodal unlabeled data. The MDFCL pre-training pipeline shown in Fig. 1(a) comprises three components: a molecular graph data encoding **(GDE)** module, a molecular sequence data encoding **(SDE)** module, and a modal integration and graph contrasting **(MIGC)** module. The MIGC module constructs the training objective loss based on latent representations from positive augmentation molecule pairs and negative pairs. The molecular adaptive augmentation strategies shown in Fig. 1(b) are designed to construct augmentation instances to satisfy the pre-training condition. These augmentation instances include side-chain repetition, backbone disruption, and side-chain generation. In Fig. 1(c), the pre-trained model is fine-tuned for molecular property prediction.

3.3. Graph data encoding module

3.3.1. Adaptive augmentation strategies

Molecular adaptive augmentation strategies in the GDE module are designed to operate on the molecular backbones and side chains. This ensures the legitimacy of generating augmentation samples and adaptive exploration of molecular structure characteristics. First, we set the following definitions:

Definition 1 (*Constructing Molecular Backbone*)**.** For a molecule G, we repeatedly delete any nodes with a degree of 1. When the degree of each node exceeds 1, the molecular backbone graph G_{bb} is obtained.

Definition 2 (*Constructing Molecular Side Chains*)**.** For a molecule G and its backbone graph G_{bb}, we obtain the subgraph $\tilde{G} = G - G_{bb}$. Subsequently, we obtain the side chain set $SideC_G = \{SC_G^1, SC_G^2, \ldots, SC_G^n\}$ of G by removing the isolated nodes in $\tilde{G}$. Furthermore, we define a fixed-length queue $SC = \{SC^1, SC^2, \ldots, SC^{500}\}$ of length 500 to record the discovered side chains of molecules in the dataset.

Specifically, given a molecular graph $G = (V, E)$, the k-core algorithm is used to obtain the backbone $\mathcal{G}_{bb}$ as

$$\mathcal{G}_{bb} = \mathrm{CORE}(G), s.t. degree(v) \geq 2, \forall v \in G, \tag{3}$$

where $\mathrm{CORE}(\cdot)$ represents the extraction of subgraphs with a node degree of 2 or higher. Subsequently, the set of side chains *SideC* is obtained by computing the difference set of graphs $\mathcal{G}_{bb}$ and G. Based on the molecular backbones and side chains, we then develop adaptive augmentation strategies to generate four types of augmentation instances, as shown in Fig. 1(b).

The generation process of all augmentation instances adopts a similar mechanism, except for differences in the backbones and side chains. First, we calculate the implicit chemical valence (IV) for each potential linkage site in the backbone. The IV value reflects the number of chemical bonds formed by the atoms, thereby determining molecular topology and connectivity. We use $IV_G = \{IV_G^1, IV_G^2, \ldots, IV_G^n\}$ to represent all connection sites set in molecule G with IV values exceeding 1. We then sample a side chain from *SideC* and assemble it to a linkage position in IV_G to generate a novel molecule. Subsequently, a series of chemical corrections – such as overlapping bonds, charge states, and isolated atoms – is performed on the generated molecule to ensure its validity and legality. The augmentation strategy is inspired by modern pharmaceutical engineering, thus providing a more accurate and fine-grained characterization of the chemical latent space. Specifically, four types of augmentation samples are generated as follows:

(1) Hard positive samples. Given a molecular graph G, a side chain is randomly sampled from its corresponding side chain set $SideC_G$ and subsequently assembled to a random linkage position in the IV_G to construct hard positive samples G_{hp}. As shown in Fig. 1(b), the upper-left subplot depicts the repeated addition of the red side chain. Because the molecules formed by adding existing side chains are similar to the original molecules, they are defined as hard positive samples.

(2) Soft positive samples. Soft positive samples G_{sp} are constructed by two methods: introducing novel side chains and adjusting the positions of existing side chains. The former method is deployed when side

chains must be sampled from $\{SC-SideC_G\}$, whereas the latter method is used to re-select a new linkage position from IV_G for the exiting side chains. In the bottom-left subplot of Fig. 1(b), the side chain marked in red is interrupted, and the blue structure is the newly introduced side chain (hydroxyl group) that enhances the polarity of the molecule. Using this operation, the model can be forced to learn the properties of molecular side chains.

(3) Soft negative samples. The primary distinction between negative and positive samples lies in their different backbone structures. We randomly perturb the backbone of molecule G to generate soft negative samples G_{sn}. In the upper-right subplot of Fig. 1(b), the tietane structure in the molecule is broken, resulting in the loss of its related properties. Thus, the model can explore the chemical semantic correlation between the properties and structures of molecules by capturing the property changes caused by structural alterations.

(4) Hard negative samples. All other data except the anchor sample in a batch of inputs can be regarded as hard negative samples G_{hn}. Furthermore, the backbone and side chains can be randomly destroyed to produce G_{hn}. In the bottom-right subplot of Fig. 1(b), the anchor molecule generates hard negative samples by destroying the backbone and shuffling or removing the side chains.

3.3.2. Encoding molecular graphs via graph neural networks

Typically, GNN frameworks [5] follow the message-passing paradigm, which iteratively extracts graph features through alternating message passing and update operations. During the message-passing phase, each node aggregates neighbor information based on its neighbors and edges in the graph. Then, the update operation updates the node embedding by aggregating the node's current state and the neighbor information. Here, two prevailing GNN models – a graph convolutional network (GCN) and a graph isomorphism network (GIN) – are employed as molecular graph encoders φ_g to learn graph modality features. To extract structure information from graphs, the GIN uses weighted summation via multilayer perceptron to aggregate the node information as follows:

$$h_v^k = \mathrm{MLP}^{(k)}\left(\left(1+\epsilon^k\right)\cdot h_v^{k-1} + \sum_{u\in\mathcal{N}(v)} h_u^{k-1}\right). \tag{4}$$

In MDFCL, READOUT$(\cdot)$ utilizes summation pooling to produce a graph representation of each layer h_G^k. Next, the messages from K layers are concatenated to yield the final molecular graph modality representation h^G as follows:

$$h_G^k = \mathrm{READOUT}\left(\left\{h_v^k \mid v\in G\right\}\right), \tag{5}$$

$$h^G = \mathrm{CONCAT}\left(h_G^k \mid k=0,1,\ldots,K\right). \tag{6}$$

3.4. Sequence data encoding module

The SDE module takes molecular SMILES to incorporate sequence modal knowledge. First, the SDE module converts four types of graph augmentation samples obtained in the GDE module into corresponding SMILES sequences as sequential augmentation samples. Since the SMILES sequence of a molecule is not unique, different SMILES of the initial molecules are also used as hard-positive sample components. That is, the hard positive sequence augmentation samples consist of two parts: the distinct SMILES of the original molecules, and the hard positive graph augmentation conversion components. The remaining three types of samples are generated using the corresponding graph augmentation samples.

When the topological structural features of molecules are encoded in graph-level learning, a simplex 1D convolutional network is employed to learn molecular sequence modality features. For a molecular sequence $S = \{s_1, s_2, \ldots, s_n\}$, each atom s_i is coded by an integer according to its associated letter notation (e.g., 1 for carbon atoms (C), 2 for nitrogen atoms (N), 3 for oxygen atoms (O)). Thus, the SMILES sequence is expressed as a sequence of integers. To cover the sequence lengths of most molecules and facilitate training, the module cuts or fills the sequence into a fixed length sequence of 300, with Shorter sequences filled with zero. These integer sequences are passed as inputs into the embedding layer, which returns the initial feature representation. After being encoded by the three convolutional layers, each atom receives a representation from the last layer of the sequence encoder φ_s; that is, $\{h_1^S, h_2^S, \ldots, h_n^S\} = \varphi_s(S)$. Finally, the max-pooling function is applied to obtain the molecular sequence modality representation, i.e., $h^S = pool(h_1^S, h_2^S, \ldots, h_n^S)$.

3.5. Modal integration and graph contrasting module

Suppose $h^G, h_{hp}^G, h_{sp}^G, h_{sn}^G$, and h_{hn}^G are graph modality representations of an anchor molecule and the four corresponding types of augmentation, respectively, learned by the GDE module. Similarly, $h^S, h_{hp}^S, h_{sp}^S, h_{sn}^S$, and h_{hn}^S denote the sequential modality representations learned by the SED module. To fuse modal information, we use the concatenation operation to obtain the final representation h of the anchor molecule:

$$h = \mathrm{CONCAT}(h^G, h^S) \in \mathbb{R}^{2d}. \tag{7}$$

Following [11,17], we apply a nonlinear projection modeled by MLP networks to the representation:

$$z = \mathrm{MLP}(h) \in \mathbb{R}^d. \tag{8}$$

Similarly, the projections (i.e., z_{hp}, z_{sp}, z_{hn}, and z_{sn}) of the four types of augmentation samples can be obtained using Eqs. (7)–(8). By maximizing the similarity of anchor molecules to positive samples while minimizing their similarity to negative samples, the model can effectively learn the structural features and semantic distribution of molecules. To achieve this, we are inspired by [19] to define the following training objective function:

$$\mathcal{L} = -\log \frac{\exp\left(\mathrm{sim}\left(z, z_+\right)/\tau\right)}{\exp\left(\mathrm{sim}\left(z, z_+\right)/\tau\right) + \sum_{z_-}\exp\left(\mathrm{sim}\left(z, z_-\right)/\tau\right)}, \tag{9}$$

where z, z_+, and z_- indicate the projection representations of the anchor, positive, and negative samples, respectively, and τ is a temperature parameter. Notably, both z_+ and z_- consist of two components, i.e., z_{hp} and z_{sp}, z_{sn} and z_{hn}, respectively. These augmentation samples refine the degree of similarity to anchor samples, resulting in a more accurate characterization of the chemical space. The function $\mathrm{sim}(z_i, z_j) = \frac{z_i^{\mathrm{T}} z_j}{|z_i||z_j|}$ is the cosine similarity function, where z_i and z_j denote molecular vector representations.

Furthermore, we establish three metric losses on the basis of Eq. (9) to refine the similarity between different types of samples.

(1) Anchor and hard positive loss. First, we calculate the similarity between anchor samples and hard negative samples, as well as that between anchor samples and soft negative samples as follows:

$$S_{z,z_{sn}} = \sum_{z_{sn}: z\in\mathcal{Z}, z_{sn}\in\mathcal{Z}_{sn}} \exp\left(\mathrm{sim}\left(z, z_{sn}\right)/\tau\right), \tag{10}$$

$$S_{z,z_{hn}} = \sum_{z_{hn}: z\in\mathcal{Z}, z_{hn}\in\mathcal{Z}_{hn}} \exp\left(\mathrm{sim}\left(z, z_{hn}\right)/\tau\right), \tag{11}$$

where $\mathcal{Z}$, $\mathcal{Z}_{sn}$, and $\mathcal{Z}_{hn}$ are the sets of anchor, soft negative, and hard negative samples, respectively, in the batch input data. The similarity between anchor samples and hard positive samples is calculated by

$$\mathcal{R}_{z,z_{hp}} = \sum_{q: z\in\mathcal{Z}, q\in\mathcal{Z}\cup\mathcal{Z}_{hp}} \mathbb{1}(z,q)\exp\left(\mathrm{sim}\left(z,q\right)/\tau\right), \tag{12}$$

where the indicator function $\mathbb{1}(\cdot) = 0$ when $z = q$, and 1 otherwise. $\mathcal{Z}_{hp}$ is the set of all hard positive samples in the input batch. The indicator function prevents computational similarity between each anchor sample and itself. Based on (10)–(12), the loss function of the anchor and hard positive samples is finally formalized as:

$$\mathcal{L}_{z,z_{hp}} = -\log \frac{\exp\left(\mathrm{sim}\left(z, z_{hp}\right)/\tau\right)}{\mathcal{R}_{z,z_{hp}} + \alpha S_{z,z_{sn}} + \beta S_{z,z_{hn}}}, \tag{13}$$

Table 2
Statistics of the datasets.

Task type	Classification							Regression					
Datasets	BBBP	Tox21	ClinTox	HIV	BACE	SIDER	MUV	FreeSolv	ESOL	Lipo	QM7	QM8	QM9
Num. of Mol.	2039	7831	1478	41 127	1513	1427	93 087	642	1128	4200	6830	21 786	130 829
Num. of tasks	1	12	2	1	1	27	17	1	1	1	1	12	8

where α and β are weight hyperparameters that control the loss values of the negative samples.

(2) Anchor and soft positive loss. Similar to the above-mentioned loss calculation process, the loss based on anchor and soft positive samples can be calculated by the following formula:

$$\mathcal{R}_{z,z_{sp}} = \sum_{q:\, z\in\mathcal{Z},q\in\mathcal{Z}\cup\mathcal{Z}_{sp}} \mathbb{1}(z\neq q)\exp\left(\mathrm{sim}\left(z,q\right)/\tau\right), \tag{14}$$

$$\mathcal{L}_{z,z_{sp}} = -\log\frac{\exp\left(\mathrm{sim}\left(z,z_{sp}\right)/\tau\right)}{\mathcal{R}_{z,z_{sp}} + \alpha S_{z,z_{sn}} + \beta S_{z,z_{hn}}}. \tag{15}$$

(3) Hard positive and soft positive loss. Since the hard and soft positive samples are both positive samples with respect to the anchor samples, there is a certain similarity between the two. Accordingly, we organize hard and soft positive samples into positive sample pairs and construct the following contrastive loss:

$$\mathcal{R}_{z_{hp},z_{sp}} = \sum_{\substack{q:\, z_{hp}\in\mathcal{Z}_{hp},\\ q\in\mathcal{Z}_{hp}\cup\mathcal{Z}_{sp}}} \mathbb{1}(z_{hp}\neq q)\exp\left(\mathrm{sim}\left(z_{hp},q\right)/\tau\right), \tag{16}$$

$$\mathcal{L}_{z_{hp},z_{sp}} = -\log\frac{\exp\left(\mathrm{sim}\left(z_{hp},z_{sp}\right)/\tau\right)}{\mathcal{R}_{z_{hp},z_{sp}} + \alpha S_{z,z_{sn}} + \beta S_{z,z_{hn}}}. \tag{17}$$

Finally, the three losses are summed with variable weights as the overall training objective:

$$\mathcal{L} = \mathcal{L}_{z,z_{hp}} + \lambda\mathcal{L}_{z,z_{sp}} + \mu\mathcal{L}_{z_{hp},z_{sp}}, \tag{18}$$

where λ and μ are hyperparameters used to balance the losses. After pre-training, we can fine-tune the molecular multimodal encoder φ for property prediction tasks.

4. Experiments

4.1. Settings

A server with an Intel(R) Core(TM) i9-10980XE CPU @ 3.00 GHz, 128G memory, and two NVIDIA GeForce GTX 3090 GPUs with 48 GB of memory is used to conduct the experiments. Our model is deployed on the Ubuntu platform with Pytorch 1.8.2. For pre-training MDFCL, we used approximately 10M unlabeled molecules selected from the PubChem database. The pre-training dataset is randomly divided into training and validation sets according to a 95/5 ratio. The batch size is set to 512 and average pooling is used. The loss function is adjusted via the Adam optimizer with a learning rate of 0.001 and weight decay of 1e−5.

For downstream tasks fine-tuning, 13 molecular datasets from MoleculeNet are utilized, as summarized in Table 2. For a fair comparison, following previous work [11], these datasets are split into training/validation/test sets according to 80/10/10. We optimize the parameters using cross-entropy loss and ℓ_1 loss for classification and regression, respectively. We select the ROC-AUC as a metric to verify classification tasks [11,17]. The RMSE metric is selected to evaluate the FreeSolv, ESOL, and Lipo datasets, and MAE is selected for the QM7, QM8, and QM9 datasets. Reported results are derived from the mean and standard deviation of three individual runs on the test set. Several representative molecular supervised learning methods are adopted for comparison, including SchNet [31], D-MPNN [32], and MGCN [33]. We also employ several classical and advanced self-supervised methods – such as HierMRL [17], MolCLR [11], N-Gram [18], and Pre-GNN [34] – as additional baselines.

4.2. Molecular property prediction results

We evaluate the performance of MDFCL on the molecular property prediction tasks. The results for the classification and regression datasets are presented in Tables 3 and 4, respectively. As shown in Table 3, we make the following observations: (1) Compared with the supervised learning models, MDFCL achieves the best results for all datasets except for BBBP. For the BBBP dataset, MDFCL is slightly weaker because the additional information introduced by SchNet and MGCN (atomic interaction and molecular conformation information) is highly correlated with the domain of the BBBP dataset. In contrast, our model significantly outperforms the traditional GCN, GIN, and D-MPNN models. (2) Compared with the self-supervised model, MDFCL achieves the best performance in all datasets except for the BBBP and HIV datasets. Specifically, the N-Gram method utilizes random walks of length n to obtain node embeddings and then combines them to constitute graph embeddings. This method can randomly capture the essential parts of a molecular graph, resulting in better performance.

The regression results represented in Table 4 indicate that MDFCL significantly outperforms the original GCN and GIN model. Except for the QM9 dataset, wherein SchNet and MGCN could capture the atomic interaction features, MDFCL achieves competitive performance for all datasets. In addition, the results of MDFCL are mixed compared to those of D-MPNN, which is a specific supervised model more beneficial for regression tasks. Compared to the self-supervised models, MDFCL achieves higher performance on all six datasets. The significant results for the six datasets suggest that molecular representation learning with multimodal information fusion significantly improves model performance.

4.3. Comparison of augmentation strategies

We compare the proposed adaptive augmentation strategies with several traditional techniques, including the masking of nodes, edges, and subgraphs. The left subplot of Fig. 2 shows the ROC-AUC measures, with standard deviations, obtained for the seven classification datasets. Overall, the traditional augmentation algorithms exhibit inferior performance and high standard deviations. The proposed augmentation strategies yield greater improvements on the classification tasks. Furthermore, considering some inherent substructures prevalent in molecular graphs, the subgraph masking strategy exhibits superior efficacy compared with that of node masking. The right subplot of Fig. 2 shows the RMSE and MAE scores, with standard deviations, obtained for the six regression datasets. The proposed strategies exhibit excellent performance and low standard deviation on the ESOL and Lipo datasets. Interestingly, the subgraph masking gains some superiority on the QM8, QM9, and FreeSolv datasets and is comparable to our methods on the QM7 dataset. The likely reason is that the molecular structures in these datasets are very simple so that it cannot provide more valuable information for our model focusing on backbones and side-chains.

We further investigate the effects of four data augmentation methods (None, Node&Edge Masking, Subgraph Masking, and Ours) on the molecular distribution. Specifically, we visualize learned molecular representations in the latent space with and without performing pre-training operations on the BBBP dataset. From left to right, Fig. 3 visualizes the visualization results of the molecular distribution under different data augmentation strategies. The application of augmentation strategies balances the data distribution and facilitates the

Table 3
The mean and standard deviation of the ROC-AUC (%) ↑ on classification benchmarks are reported. The optimal performance within each dataset is underlined, and the most favorable outcomes associated with the supervised (SL) and self-supervised (SSL) paradigms are bolded.

	Dataset	BBBP	Tox21	ClinTox	HIV	BACE	SIDER	MUV
	#Molecules	2039	7831	1478	41 127	1513	1427	93 087
	#Tasks	1	12	2	1	1	27	17
SL	GCN	71.8 ± 0.9	70.9 ± 2.6	62.5 ± 2.8	74.0 ± 3.0	71.6 ± 2.0	53.6 ± 3.2	71.6 ± 4.0
SL	GIN	65.8 ± 4.5	74.0 ± 0.8	58.0 ± 4.4	**75.3 ± 1.9**	70.1 ± 5.4	57.3 ± 1.6	71.8 ± 2.5
SL	SchNet	84.8 ± 2.2	**77.2 ± 2.3**	71.5 ± 3.7	70.2 ± 3.4	76.6 ± 1.1	53.9 ± 3.7	71.3 ± 3.0
SL	MGCN	**85.0 ± 6.4**	70.7 ± 1.6	63.4 ± 4.2	73.8 ± 1.6	73.4 ± 3.0	55.2 ± 1.8	70.2 ± 3.4
SL	D-MPNN	71.2 ± 3.8	68.9 ± 1.3	**90.5 ± 5.3**	75.0 ± 2.1	**85.3 ± 5.3**	**63.2 ± 2.3**	**76.2 ± 2.8**
SSL	Pre-GNN	70.8 ± 1.5	78.7 ± 0.4	78.9 ± 2.4	80.2 ± 0.9	85.9 ± 0.8	65.2 ± 0.9	81.4 ± 2.0
SSL	N-Gram	**91.2 ± 3.0**	76.9 ± 2.7	85.5 ± 3.7	**83.0 ± 1.3**	87.6 ± 3.5	63.2 ± 0.5	81.6 ± 1.9
SSL	$MolCLR_{GIN}$	73.6 ± 0.5	72.3 ± 0.7	88.9 ± 1.7	74.8 ± 1.1	80.5 ± 0.3	61.3 ± 1.1	73.9 ± 2.2
SSL	$HierMRL_{GIN}$	74.5 ± 1.6	79.2 ± 0.6	96.5 ± 1.5	78.2 ± 1.1	87.7 ± 1.7	68.6 ± 1.1	88.0 ± 1.0
SSL	$MDFCL_{GCN}$	73.5 ± 1.8	76.2 ± 1.4	90.8 ± 1.5	80.7 ± 1.5	**87.9 ± 0.9**	**70.0 ± 1.0**	87.4 ± 2.1
SSL	$MDFCL_{GIN}$	75.1 ± 1.7	**79.5 ± 1.0**	**96.8 ± 1.7**	78.7 ± 1.3	85.4 ± 1.2	68.7 ± 0.9	**88.6 ± 1.2**

Table 4
The performance of the different methods on six regression benchmarks. The evaluation metrics are RMSE (↓) for ESOL, FreeSolv, and Lipo, and MAE (↓) for QM7, QM8, and QM9.

	Dataset	FreeSolv	ESOL	Lipo	QM7	QM8	QM9
	#Molecules	642	1128	4200	6830	21 786	130 829
	#Tasks	1	1	1	1	12	8
SL	GCN	2.87 ± 0.14	1.43 ± 0.05	0.85 ± 0.08	122.9 ± 2.2	0.0366 ± 0.0011	5.796 ± 1.969
SL	GIN	2.76 ± 0.18	1.45 ± 0.02	0.85 ± 0.07	124.8 ± 0.7	0.0371 ± 0.0009	4.741 ± 0.912
SL	SchNet	3.22 ± 0.76	1.05 ± 0.06	0.91 ± 0.10	**74.2 ± 6.0**	0.0204 ± 0.0021	0.081 ± 0.001
SL	MGCN	3.35 ± 0.01	1.27 ± 0.15	1.11 ± 0.04	77.6 ± 4.7	0.0223 ± 0.0021	**0.050 ± 0.002**
SL	D-MPNN	**2.18 ± 0.91**	**0.98 ± 0.26**	**0.65 ± 0.05**	105.8 ± 13.2	**0.0143 ± 0.0022**	3.241 ± 0.119
SSL	Pre-GNN	2.83 ± 0.12	1.22 ± 0.02	0.74 ± 0.00	110.2 ± 6.4	0.0191 ± 0.0003	4.349 ± 0.061
SSL	N-Gram	2.51 ± 0.19	1.10 ± 0.03	0.88 ± 0.12	125.6 ± 1.5	0.0320 ± 0.0032	7.636 ± 0.027
SSL	$MolCLR_{GIN}$	2.83 ± 0.20	1.29 ± 0.01	0.69 ± 0.08	68.5 ± 2.0	0.0183 ± 0.0013	3.496 ± 0.118
SSL	$HierMRL_{GIN}$	2.48 ± 0.17	1.09 ± 0.05	0.69 ± 0.01	76.44 ± 2.8	0.0178 ± 0.0016	2.983 ± 0.030
SSL	$MDFCL_{GCN}$	2.45 ± 0.19	1.09 ± 0.04	0.73 ± 0.05	**67.07 ± 3.2**	0.0176 ± 0.0037	7.500 ± 0.051
SSL	$MDFCL_{GIN}$	**2.44 ± 0.23**	**1.05 ± 0.05**	**0.68 ± 0.03**	71.16 ± 4.1	**0.0173 ± 0.0041**	**2.653 ± 0.075**

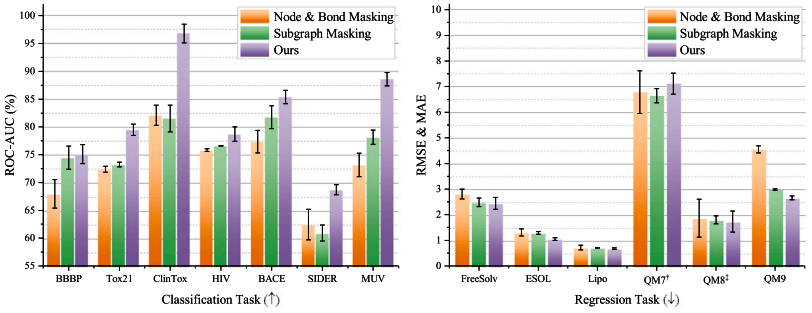

Fig. 2. Performance results of different augmentation strategies in seven classification and six regression datasets. † and ‡ denote operations (÷10) and (×100) on the indicator values, respectively.

discrimination of distinct sample classes in the latent space, irrespective of whether the model undergoes fine-tuning for the downstream task. Furthermore, the proposed data augmentation strategy effectively guides the distribution of molecules in the latent space, facilitating their clustering based on class labels. From top to bottom, Fig. 3 visualizes results with and without fine-tuning under the same augmentation strategy. Evidently, the fine-tuning operation contributes more valuable information for the model.

4.4. Ablation study

To further evaluate the sub-modules of MDFCL, we conduct an ablation study. The results are evaluated based on the average metrics for seven classification datasets and five regression datasets (w/o QM7). For the regression tasks, the metrics are averaged between two groups of experiments, and divided according to RMSE and MAE. Table 5 shows results for molecular sequence and graph modal augmentation modules. Overall, both modal augmentation modules are effective in boosting model performance. An interesting phenomenon is that graph

Table 5
Ablation experiment results of the MDFCL model. ✓ represents different combinations of graph and sequence augmentation modules.

Augmentation strategy		Task		
Graph augmentation	Sequence augmentation	Classification ROC-AUC (%)	Regression RMSE	Regression MAE
–	✓	78.94	1.532	1.394
✓	–	81.81	1.461	1.361
✓	✓	81.83	1.391	1.349

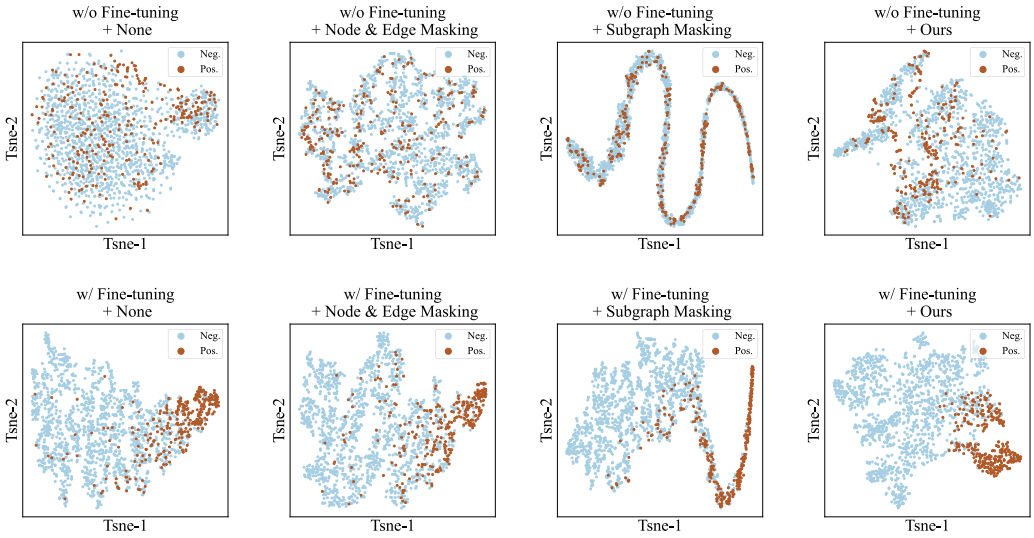

Fig. 3. Latent space visualization of BBBP dataset under different augmentation strategies with (w/) or without fine-tuning (w/o).

Table 6
Ablation experiment results of graph augmentation module. ✓ represents different combinations of augmentation instances.

Augmentation strategy			Task		
Hard positives	Soft positives	Negatives	Classification ROC-AUC (%)	Regression RMSE	Regression MAE
–	✓	✓	75.01	2.050	1.783
✓	–	✓	75.99	1.777	1.851
✓	✓	–	77.43	1.666	1.816
✓	✓	✓	81.81	1.452	1.636

augmentation significantly influences classification and regression tasks more than sequence augmentation. It is observed that removing graph augmentation leads to a −2.89% change in ROC-AUC, with RMSE and MAE changing by +0.141 and +0.045, respectively. In contrast, dropping sequence augmentation affects ROC-AUC, RMSE, and MAE by only −0.02%, +0.07, and +0.012, respectively. Furthermore, Table 6 shows the impact of different graph augmentation instances on model performance. The results demonstrate that three augmentation samples can provide valuable information and improve performance, and that the hard positive samples are the most effective. One possible explanation is that the hard positive samples more effectively direct the distribution of molecules owing to their high structural similarity to the anchors.

4.5. Investigation of MDFCL representation

We visualize the representations learned by MDFCL following dimensionality reduction by t-SNE. If molecules possess similar structures, they should be mapped to adjacent domains in the latent space. Fig. 4 depicts 50K molecules from the PubChem database embedded in 2D by t-SNE, colored with the corresponding molecular weights. As shown in Fig. 4, MDFCL can map molecules with similar backbones and functional group structures to adjacent domains in the latent space. This indicates that MDFCL can also learn the intrinsic connections between molecules from massive unlabeled data, as molecules with similar structures have similar properties. We also compare the MDFCL-learned representations with two conventional molecular fingerprints: RDKFP and ECFP. Specifically, given a query molecule, we first compute the cosine distance of its representation with all pretrained molecules. Subsequently, the nine molecules closest to the query molecule in the MDFCL representation domain are visualized and labeled with two chemical fingerprint similarity values. These molecular representations learned via MDFCL exhibit a high degree

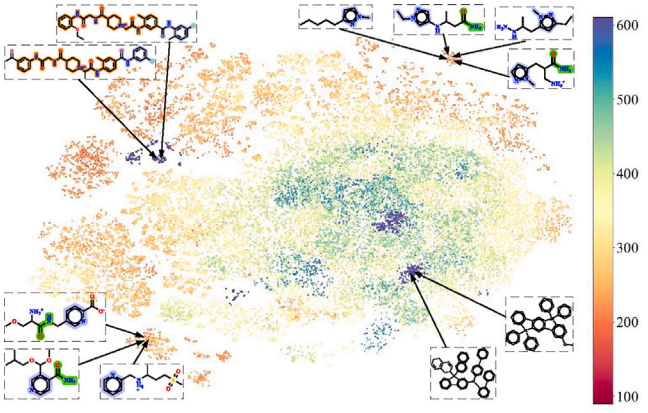

Fig. 4. Molecular representations learned by MDFCL are visualized after dimensionality reduction by t-SNE. There are 50k molecules in this Figure. Data points are colored with their corresponding molecular weights. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of concordance with conventional chemical FPs. In Fig. 5(a), RDKFP ranges from 0.857 to 0.961, and the closest molecule is almost identical to the first query molecule, except for one bromine atom. In addition, we can observe that these selected molecules possess the same functional groups and backbone structures, whereas their side chains are diverse, such as bromobenzene, thiophene, and formaldehyde. A similar phenomenon is observed for the second query molecule, shown in Fig. 5(b). This indicates that MDFCL can rationally incorporate chemical domain knowledge and autonomously learn efficient representations of molecules.

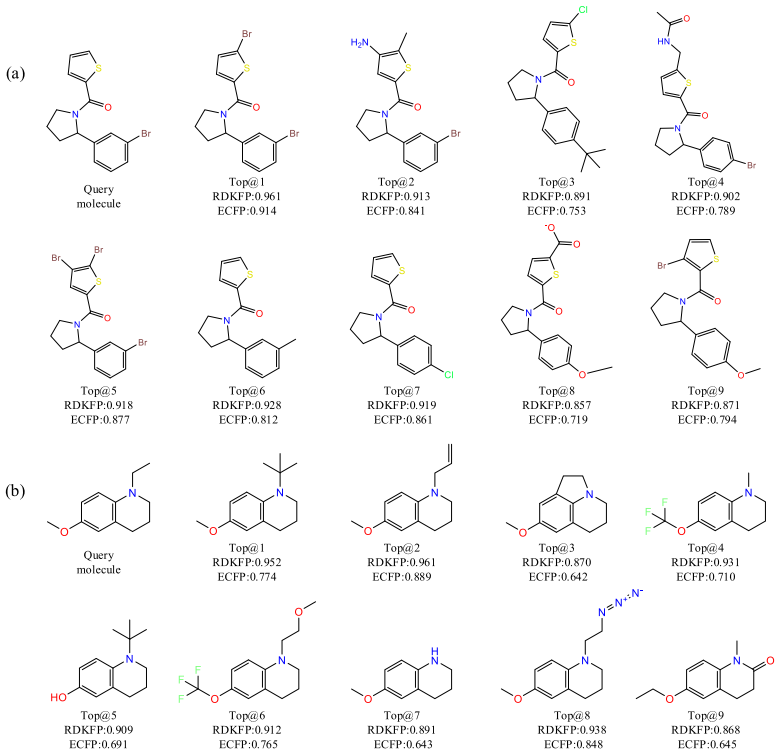

Fig. 5. Comparison of MDFCL-learned molecular representations and conventional fingerprints. The two query molecules (PubChem ID61061637 and PubChem ID70645498) and the corresponding nine closest molecules in the MDFCL representation domain labeled with RDKFP and ECFP similarities values are visualized.

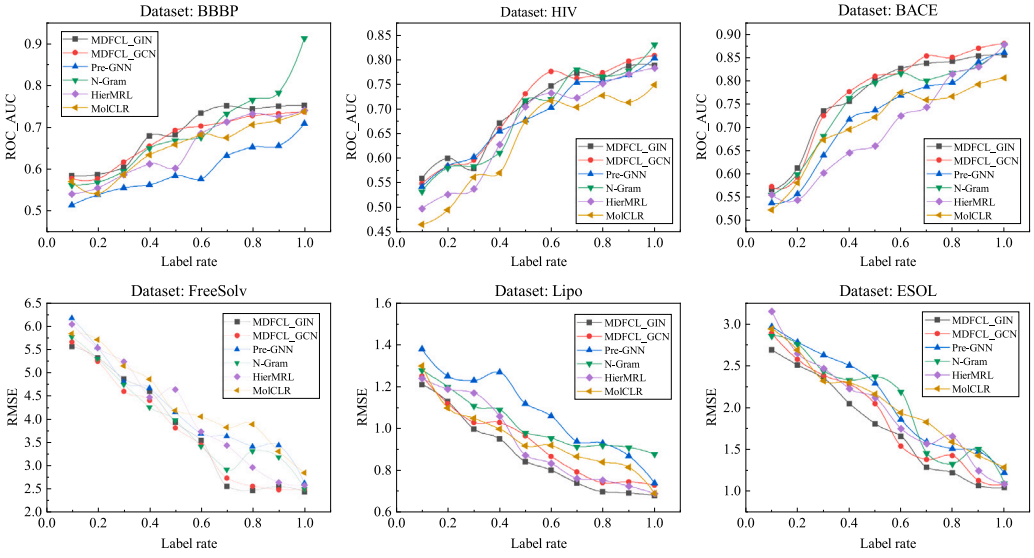

Fig. 6. The results of several self-supervised methods with different fine-tuning label rates.

4.6. Fine-tuning with different label rates

For further comparison, we vary the fine-tuning label rate between {0.1, 0.2, . . . , 1} to simulate a realistic self-supervised scenario wherein only a few labeled instances are available. When the label rate is set to 1, it indicates that all established labeled samples are utilized for fine-tuning, similar settings to Tables 3 and 4. The corresponding average ROC-AUC and RMSE values of several self-supervised methods are shown in Fig. 6. Overall, each method exhibits substantially improved performance with an increase in the fine-tuning label rate. This phenomenon is attributed to the fact that models can learn valuable information from more labeled instances. On the HIV and BACE datasets, MDFCL achieves competitive performance compared to other self-supervised methods under most of the labeling rate settings. On the BBBP dataset, the N-Gram method slightly outperforms MDFCL when fine-tuning label rates exceed 0.8. For the regression tasks across the FreeSolv, Lipo, and ESOL datasets, MDFCL demonstrates superior performance under most fine-tuning label rates. Notably, MDFCL is exceptionally competitive at lower fine-tuning label rates. For instance, at a label rate of 0.1, MDFCL achieves RMSE values of 1.21 and 2.69 for

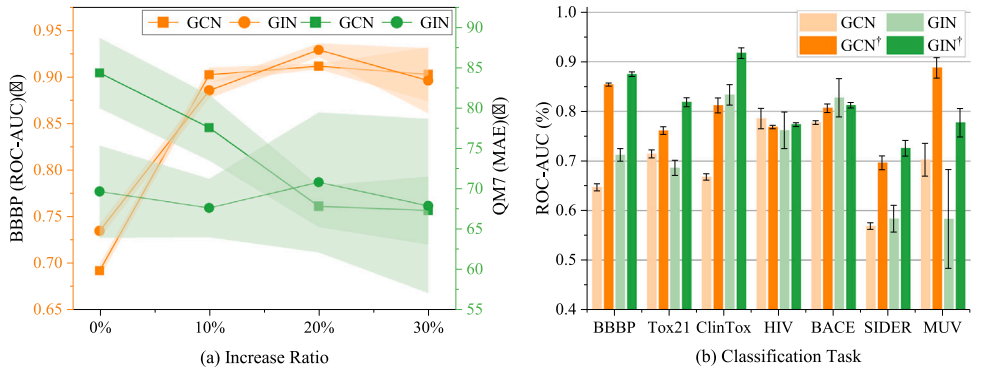

Fig. 7. Results of the supervised learning models after introducing augmentation samples. The superscript † indicates 10% augmentation samples.

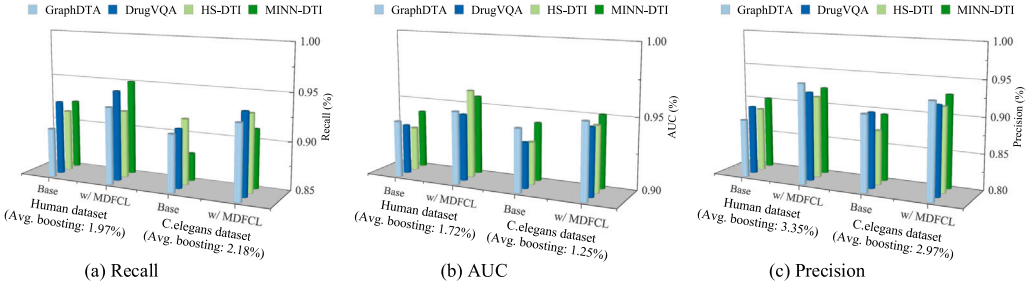

Fig. 8. MDFCL boosts the DTI prediction performance on Human and C.elegans datasets.

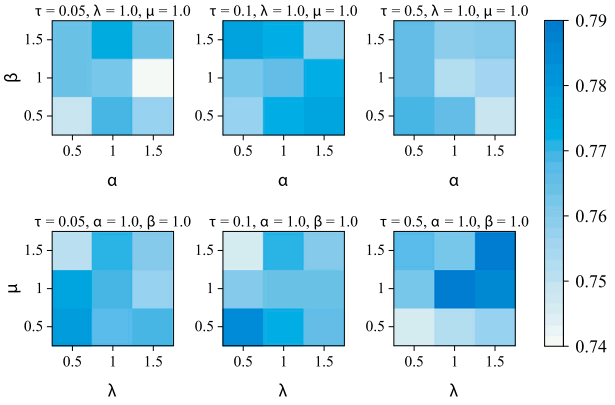

Fig. 9. ROC-AUC (%) metrics for different combinations of hyperparameters in the BBBP dataset.

the Lipo and ESOL datasets, respectively, significantly outperforming the comparison models.

4.7. Augmentation strategies under scare sample conditions

We investigate the impact of different quantities of augmentation samples by deploying the GCN and GIN in an environment with few samples. Fig. 7(a) presents results for the BBBP and QM7 datasets with different ratios of added augmentation samples. Overall, model performance improves significantly as the ratio increases from 0 to 10%. As more instances are introduced, the standard deviation and instability increase along with the potential for label errors. Subsequently, 10% augmentation samples are introduced into the experiment. The results shown in Fig. 7(b) indicate that introducing a small number of augmentation samples into a small dataset generally has a positive effect. Furthermore, unlabeled augmentation samples may stabilize the model's performance.

4.8. Boosting performance of drug-target interaction prediction

We extend MDFCL as a universal plugin for molecular representation learning and embed it into drug-target interaction prediction (DTI) to further explore its generalizability. Specifically, four representative DTI models are selected as base models, namely GraphDTA [12], DrugVQA [35], HS-DTI [25], and MINN-DTII [36]. We replace the drug molecule encoder components in the base models with the pre-trained MDFCL (w/ MDFCL) while maintaining all other experimental procedures to ensure a fair comparison for the Human and C.elegans datasets. As shown in Fig. 8, MDFCL significantly boosts the performance of the base models, with a 3.35% average precision improvement on the Human dataset and 2.97% average precision improvement on the C.elegans dataset. The performance improvements are also evident in terms of AUC and Recall, reconfirming the potent capability of MDFCL in molecular representation learning.

4.9. Parameter sensitivity experiments

We further examine the effects of hyperparameters. Fig. 9 shows experimental results for the BBBP dataset. The three plots in the first row show the effects of different combinations of temperature parameter τ, negative instance weights α, β on the performance of loss balance parameters λ and μ, where $\tau = 0.1$, $\alpha = 0.5$, and $\beta = 1.5$ are the optimal hyperparameter values. The three plots in the second row visualize the impact of different combinations of τ, λ, and μ on performance under α and β conditions, where $\tau = 0.5$, $\lambda = 1$, and $\mu = 1$ are the optimal hyperparameter values. In accordance with these results, we generalize the hyperparameter combination $\tau = 0.5$, $\lambda = 1$, $\mu = 1$, $\alpha = 0.5$, and $\beta = 1.5$ for all experiments.

5. Conclusion

To resolve the issues inherent to scarce labeled data and unimodal data when characterizing molecular information, we propose the MDFCL framework for molecular property prediction. The underlying concept of MDFCL encompasses four components. First, new adaptive augmentation strategies are designed to generate four types of molecular augmentation instances. For graph modal data, adaptive augmentation strategies – including backbone perturbation, side-chain generation, and side-chain deletion – are constructed on the molecular backbones and side chains. Second, the modal switching module transforms generated graph augmentation instances into sequence data. Third, using the augmentation strategies, a graph contrastive learning framework is constructed for pre-training by contrasting positive molecule pairs against negative molecular pairs. Finally, MDFCL is subsequently fine-tuned on the downstream property prediction benchmarks. By integrating multimodal molecular data, MDFCL significantly improves performance on both classification and regression tasks compared to supervised approaches. Benefiting from the proposed adaptive augmentation strategies, MDFCL surpasses other self-supervised methods in most molecular benchmarks and can steer the distribution of molecular representations in the latent chemical space.

In future studies, we will explore the multimodal learning of molecules in greater depth. Since MDFCL might not be suitable for macromolecules (e.g., proteins), we intend to make it suitable for macromolecular pre-training. Furthermore, we intend to focus on a wider range of applications for MDFCL, including drug discovery and materials science.

CRediT authorship contribution statement

Xu Gong: Writing – review & editing, Writing – original draft, Validation, Software, Methodology. **Maotao Liu:** Validation, Software. **Qun Liu:** Writing – review & editing, Writing – original draft, Supervision. **Yike Guo:** Validation, Supervision, Project administration. **Guoyin Wang:** Visualization, Supervision, Project administration, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

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

Acknowledgments

This work is supported by the key cooperation project of the Chongqing Municipal Education Commission (HZ2021008), partly funded by the State Key Program of National Nature Science Foundation of China (61936001), and the Doctoral Innovation Talent Program of Chongqing University of Posts and Telecommunications (No. BYJS202301).

Data availability

Data will be made available on request.

References

[1] M. Liu, Y. Yang, Q. Liu, L. Liu, G. Wang, A knowledge-driven self-supervised approach for molecular generation, IEEE/ ACM Trans. Comput. Biology Bioinform. (2024).
[2] S. Bakkali, Z. Ming, M. Coustaty, M. Rusiñol, O.R. Terrades, VLCDoC: Vision-language contrastive pre-training model for cross-modal document classification, 2022, arXiv preprint arXiv:2205.12029.
[3] J. Xia, Y. Zhu, Y. Du, Y. Liu, S. Li, A Systematic Survey of Chemical Pre-trained Models, IJCAI, 2023.
[4] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings, OpenReview.net, 2017.
[5] L. Xu, J. Peng, X. Jiang, E. Chen, B. Luo, Graph neural network based on graph kernel: A survey, Pattern Recognit. (2024) 111307.
[6] W. Hamilton, Z. Ying, J. Leskovec, Inductive representation learning on large graphs, Adv. Neural Inf. Process. Syst. 30 (2017).
[7] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, et al., Graph attention networks, Stat 1050 (20) (2017) 10–48550.
[8] Q. Chang, X. Li, Z. Duan, Graph global attention network with memory: A deep learning approach for fake news detection, Neural Netw. 172 (2024) 106115.
[9] X. Gong, Q. Liu, R. Han, Y. Guo, G. Wang, MIFS: An adaptive multipath information fused self-supervised framework for drug discovery, Neural Netw. 184 (2025) 107088.
[10] J. Xia, C. Zhao, B. Hu, Z. Gao, C. Tan, Y. Liu, S. Li, S.Z. Li, Mole-bert: Rethinking pre-training graph neural networks for molecules, in: The Eleventh International Conference on Learning Representations, 2022.
[11] Y. Wang, J. Wang, Z. Cao, A. Barati Farimani, Molecular contrastive learning of representations via graph neural networks, Nat. Mach. Intell. 4 (3) (2022) 279–287.
[12] T. Nguyen, H. Le, T.P. Quinn, T. Nguyen, T.D. Le, S. Venkatesh, GraphDTA: Predicting drug–target binding affinity with graph neural networks, Bioinformatics 37 (8) (2021) 1140–1147.
[13] J. Ross, B. Belgodere, V. Chenthamarakshan, I. Padhi, Y. Mroueh, P. Das, Large-scale chemical language representations capture molecular structure and properties, Nat. Mach. Intell. 4 (12) (2022) 1256–1264.
[14] H. Li, D. Zhao, J. Zeng, KPGT: Knowledge-guided pre-training of graph transformer for molecular property prediction, in: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022, pp. 857–867.
[15] Z. Zeng, Y. Yao, Z. Liu, M. Sun, A deep-learning system bridging molecule structure and biomedical text with comprehension comparable to human professionals, Nat. Commun. 13 (1) (2022) 862.
[16] F.-Y. Sun, J. Hoffman, V. Verma, J. Tang, InfoGraph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization, in: International Conference on Learning Representations, 2020.
[17] M. Liu, Y. Yang, X. Gong, L. Liu, Q. Liu, HierMRL: Hierarchical structure-aware molecular representation learning for property prediction, in: IEEE International Conference on Bioinformatics and Biomedicine, BIBM, 2022, pp. 386–389.
[18] S. Liu, M.F. Demirel, Y. Liang, N-gram graph: Simple unsupervised representation for graphs, with applications to molecules, in: Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS, 2019, pp. 8464–8476.
[19] Y. Wang, R. Magar, C. Liang, A. Barati Farimani, Improving molecular contrastive learning via faulty negative mitigation and decomposed fragment contrast, J. Chem. Inf. Model. 62 (11) (2022) 2713–2725.
[20] S. Han, H. Fu, Y. Wu, G. Zhao, Z. Song, F. Huang, Z. Zhang, S. Liu, W. Zhang, HimGNN: A novel hierarchical molecular graph representation learning framework for property prediction, Brief. Bioinform. 24 (5) (2023) bbad305.
[21] C. Shen, J. Luo, K. Xia, Molecular geometric deep learning, Cell Rep. Methods 3 (11) (2023).
[22] X.-b. Ye, Q. Guan, W. Luo, L. Fang, Z.-R. Lai, J. Wang, Molecular substructure graph attention network for molecular property identification in drug discovery, Pattern Recognit. 128 (2022) 108659.
[23] X. Zang, X. Zhao, B. Tang, Hierarchical molecular graph self-supervised learning for property prediction, Commun. Chem. 6 (1) (2023) 34.
[24] X. Ai, C. Sun, Z. Zhang, E. Hancock, Two-level graph neural network, IEEE Trans. Neural Networks Learn. Syst. (2022).
[25] X. Gong, M. Liu, H. Sun, M. Li, Q. Liu, HS-DTI: Drug-target interaction prediction based on hierarchical networks and multi-order sequence effect, in: 2022 IEEE International Conference on Bioinformatics and Biomedicine, BIBM, IEEE, 2022, pp. 322–327.
[26] D. Zhang, S. Xia, Y. Zhang, Accurate prediction of aqueous free solvation energies using 3d atomic feature-based graph neural network with transfer learning, J. Chem. Inf. Model. 62 (8) (2022) 1840–1848.
[27] R. Irwin, S. Dimitriadis, J. He, E.J. Bjerrum, Chemformer: A pre-trained transformer for computational chemistry, Mach. Learn. Sci. Technol. 3 (1) (2022) 15022.

[28] X. Zhang, C. Wu, Z. Yang, Z. Wu, J. Yi, C. Hsieh, T. Hou, D. Cao, MG-BERT: Leveraging unsupervised atomic representation learning for molecular property prediction, Briefings Bioinform. 22 (6) (2021).
[29] Y. Zhu, D. Chen, Y. Du, Y. Wang, Q. Liu, S. Wu, Featurizations matter: A multiview contrastive learning approach to molecular pretraining, in: ICML 2022 2nd AI for Science Workshop, 2022.
[30] H. Zhang, J. Wu, S. Liu, S. Han, A pre-trained multi-representation fusion network for molecular property prediction, Inf. Fusion 103 (2024) 102092.
[31] K.T. Schütt, H.E. Sauceda, P.J. Kindermans, A. Tkatchenko, K.R. Müller, SchNet – A deep learning architecture for molecules and materials, J. Chem. Phys. 148 (24) (2018) 241722.
[32] K. Yang, K. Swanson, W. Jin, C.W. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea, A. Palmer, V. Settels, T.S. Jaakkola, K.F. Jensen, R. Barzilay, Analyzing learned molecular representations for property prediction, J. Chem. Inf. Model. 59 (8) (2019) 3370–3388.
[33] C. Lu, Q. Liu, C. Wang, Z. Huang, P. Lin, L. He, Molecular property prediction: A multilevel quantum interactions modeling perspective, in: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI, 2019, pp. 1052–1060.
[34] W. Hu, B. Liu, J. Gomes, M. Zitnik, P. Liang, V.S. Pande, J. Leskovec, Strategies for pre-training graph neural networks, in: 8th International Conference on Learning Representations, ICLR, 2020.
[35] S. Zheng, Y. Li, S. Chen, J. Xu, Y. Yang, Predicting drug–protein interaction using quasi-visual question answering system, Nat. Mach. Intell. 2 (2) (2020) 134–140.
[36] F. Li, Z. Zhang, J. Guan, S. Zhou, Effective drug–target interaction prediction with mutual interaction neural network, Bioinformatics 38 (14) (2022) 3582–3589.

Xu Gong received the B.S., and M.S. degrees from Chongqing Normal University in China in 2017, and 2021, respectively. He is currently a Ph.D. student at College of Computer Science and Technology, Chongqing University of Posts and Telecommunications. His research interests includes causal inference, explainable graph neural network techniques, and drug design.

Maotao Liu received his B.S. degree from Chongqing University in China in 2015. He is currently working toward the M.S. degree in the Chongqing University of Posts and Telecommunications. His research focuses on the development and application of deep learning methods for knowledge driven drug discovery.

Qun Liu received her B.S. degree from Xi'An Jiaotong University in China in 1991, and the M.S. degree from Wuhan University in China in 2002, and the Ph.D from Chongqing University in China in 2008. She is currently a Professor with Chongqing University of Posts and Telecommunications. Her current research interests include complex and intelligent systems, neural networks and intelligent information processing.

Yike Guo is the vice president of Hong Kong University of Science and Technology and professor at Imperial College London. He is an IEEE fellow, fellow of the Royal Academy of Engineering (FREng), member of the Academia Europaea (MAE), fellow of the British Computer Society and a trustee of the Royal Institution of Great Britain. Professor Guo has published over 200 articles, papers, and reports. His current research interests focus on data mining, machine learning, and dig data of science.

Guoyin Wang (SM'03) received the B.S., M.S., and Ph.D. degrees from Xi'an Jiaotong University, Xian, China, in 1992, 1994, and 1996, respectively. He was at the University of North Texas, and the University of Regina, Canada, as a visiting scholar during 19981999. Since 1996, he has been at the Chongqing University of Posts and Telecommunications, where he is currently a professor, the director of the Chongqing Key Laboratory of Computational Intelligence, the Vice President of the University and the dean of the School of Graduate. He was appointed as the director of the Institute of Electronic Information Technology, Chongqing Institute of Green and Intelligent Technology, CAS, China, in 2011. He is the author of over 10 books, the editor of dozens of proceedings of international and national conferences, and has more than 300 reviewed research publications. His research interests include rough sets, granular computing, knowledge technology, data mining, neural network, and cognitive computing, etc. Dr. Wang was the President of International Rough Set Society (IRSS) 20142017. He is a Vice President of the Chinese Association for Artificial Intelligence (CAAI), and a council member of the China Computer Federation (CCF).