Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-25)
Special Track on AI4Tech: AI Enabling Critical Technologies

MolHFCNet: Enhancing Molecular Graph Representations with Hierarchical Feature Combining and Hybrid Pretraining

Duy-Long Nguyen[1] , **Duc-Luong Ho-Viet**[1] , **Anh-Thu Ngo-Tran**[1] , **Quang H. Nguyen**[1] and **Binh P. Nguyen**[2,*]

[1] Hanoi University of Science and Technology, Hanoi, 100000, Vietnam
[2] Victoria University of Wellington, Wellington, 6012, New Zealand
long.nd242076m@sis.hust.edu.vn, luonghvd@soict.hust.edu.vn, thu.nta242075m@sis.hust.edu.vn, quangnh@soict.hust.edu.vn, binh.p.nguyen@vuw.ac.nz

Abstract

Efficient molecular property prediction is crucial in bioinformatics and cheminformatics, with applications in drug discovery, materials science, and chemical engineering. This paper introduces MolHFCNet, a graph neural network designed to enhance molecular representation learning. At its core, the n-Hierarchical Features Combining (n-HFC) module aggregates information across multiple hierarchical feature spaces, effectively capturing both local and global graph structures. Unlike conventional models, n-HFC maintains computational complexity comparable to a single full-dimensional graph layer while supporting either 2D or 3D molecular graphs, ensuring flexibility across tasks. Furthermore, we propose a novel graph pretraining strategy that integrates predictive and contrastive learning, enabling the model to capture local chemical interactions and global molecular contexts for robust embeddings. Experimental results on benchmark datasets demonstrate MolHFCNet's superior accuracy and efficiency compared to state-of-the-art methods, highlighting the potential of high-order hierarchical feature learning for advancing molecular graph analysis. Our code is available at https://github.com/ndlongvn/MolHFCNet.

1 Introduction

Predicting molecular properties is crucial for drug discovery and materials science, enabling the identification of specific molecules while reducing experimental costs [Yang *et al.*, 2019]. Recent advancements in deep learning (DL), which have revolutionized fields like natural language processing (NLP), computer vision (CV), and graph analysis, prompted chemists to adopt DL techniques for molecular property prediction, particularly in chemical modeling and drug discovery [Gilmer *et al.*, 2017]. Molecular representations like SMILES and SELFIES are foundational in property prediction tasks, enabling machine learning (ML) and DL models to process molecular structures effectively. SMILES encodes molecules as sequences, making it well-suited for NLP techniques. Models like ChemBERTa [Chithrananda *et al.*, 2020] and SMILES-BERT [Wang *et al.*, 2019] used Transformers to predict molecular properties. Zhu *et al.* [2023] introduced DVMP, combining Transformers and graph neural networks (GNNs) to leverage both SMILES and molecular graphs for improved performance. SELFIES overcomes the limitations of SMILES by ensuring 100% chemical validity through context-free grammar. This robustness has enabled advances in molecular property prediction and drug discovery. For instance, SELFormer [Yüksel *et al.*, 2023], pretrained on two million SELFIES compounds representation, surpassed SMILES-based models in adverse drug reaction predictions. SELFIES has also driven innovations in cheminformatics, including genetic algorithms with DNNs [Nigam *et al.*, 2020], curiosity-driven reinforcement learning [Thiede *et al.*, 2022], and multi-objective optimization [Alberga *et al.*, 2024]. Together, SMILES and SELFIES are essential for molecular properties prediction with unique strengths.

Besides SMILES and SELFIES, 2D molecular representations based on graphs, where atoms are modeled as nodes and bonds as edges, offer richer relational information and have attracted increasing interest [Nguyen *et al.*, 2024]. This has driven the rapid development of GNNs, which excel at modeling graph-structured data and have demonstrated superior performance in predicting quantum mechanical [Yang *et al.*, 2019] and molecular properties [Zhu *et al.*, 2022; Zang *et al.*, 2023; Zhu *et al.*, 2023]. However, while 2D graph representations effectively capture molecular topology, they overlook crucial 3D geometric information, which plays a fundamental role in determining molecular properties. To address this limitation, recent studies have incorporated 3D structural data into molecular representations, leveraging advanced geometric learning techniques. Nguyen *et al.* [2024] introduced the multimodal model, which integrates 2D and 3D graph information with contrastive learning to enhance molecular representations. Similarly, Liu *et al.* [2022a] proposed GeoSSL-DMM, an SE(3)-invariant score matching approach that reformulates coordinate denoising as denoising pairwise atomic distances, achieving state-of-the-art results. Additionally, Zhou *et al.* [2023] developed the Uni-Mol model based on the SE(3) Transformer architecture, embedding 3D graph information to facilitate effective learning and representation of molecular conformers.

Despite significant advancements, traditional GNNs face several critical challenges that limit their effectiveness in

*Corresponding author

deep architectures. First, the practice of propagating full-dimensional features across layers leads to substantial computational complexity, resulting in high memory and processing demands that hinder the scalability on large or complex graphs [Duan *et al.*, 2022]. Second, traditional molecular graph learning models are typically designed with either a 2D-GNNs or 3D-GNNs backbone [Zhu *et al.*, 2022; Schütt *et al.*, 2017], limiting their adaptability across different molecular prediction tasks. To address these challenges, we introduce **MolHFCNet**, a novel hierarchical GNN tailored for molecular property prediction. While HorNet [Rao *et al.*, 2022] utilizes hierarchical convolutional layers for multi-scale feature extraction in computer vision, MolHFCNet independently adapts and extends these principles to molecular graphs by introducing the n-HFC module. Unlike HorNet, our n-HFC module is specifically designed for molecular graphs, enabling multi-scale message passing to capture both local and global structural patterns. By employing a hierarchical strategy, the n-HFC module expands intermediate feature dimensions, enabling deeper architectures to retain more distinct and informative node representations. This hierarchical scaling not only maintains computational efficiency, keeping n-HFC's complexity comparable to a single full-dimensional standard graph layer but also enhances the model's ability to learn adaptive, multi-scale representations, effectively addressing the limitations of traditional GNN architectures. Furthermore, we introduce a novel graph pretraining strategy that integrates both predictive and contrastive learning objectives, enabling the model to capture contexts and interactions. MolHFCNet also supports diverse GNN backbones, adapting to molecular topology when 3D graph is unavailable or leveraging spatial information.

2 Method

Notation: In graph representation, a molecule is modeled as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the nodes $\mathcal{V}$ correspond to the atoms, and the edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ represent the chemical bonds between them. Each node $v_i \in \mathcal{V}$ is associated with a feature vector $\mathbf{h}_i \in \mathbb{R}^d$, where d is the dimension of the atomic capturing properties. Each edge in $\mathcal{E}$ is characterized by an edge feature vector $\mathbf{e}_{ij} \in \mathbb{R}^{d_e}$ of d_e dimension. In addition to atomic features, the 3D spatial configuration of the molecule for capturing geometric and conformational properties is represented as $\mathbf{R} \in \mathbb{R}^{|\mathcal{V}|\times 3}$, where $R_i = (x_i, y_i, z_i)$ denotes the Cartesian coordinates of v_i.

2.1 Multi-task Self-Supervised Contrastive Pretraining

In this study, we introduce a graph pretraining strategy for molecular representation learning to effectively capture both local and global molecular features. Our approach leverages the complementary strengths of predictive and contrastive pretraining methods. By integrating these objectives with descriptor and fingerprint-based supervision, we establish a robust foundation for learning enriched representations.

Masked attribute prediction for 2D and 3D graphs

The masked attribute prediction method trains the model to infer missing node and edge features based on the remaining graph structure. This strategy aligns with prior works, such as those in [Hu *et al.*, 2020; Zhou *et al.*, 2023], and is utilized for both 2D and 3D molecular graphs. In 3D graphs, where nodes represent atoms with chemical properties and spatial coordinates, a dual-masking strategy–atomic feature masking and spatial coordinate masking–enhances feature learning. Given a batch of molecular graphs $\{\mathcal{G}^k\}_{k=1}^B$ with batch size B, for each graph, atom v_i^k is independently selected for masking with a fixed probability $p \in (0, 1)$. Its masked chemical features and coordinates are replaced with zero vectors, with $\mathcal{V}_m^k$ represents the set of masked nodes for graph $\mathcal{G}^k$. The masked node embeddings $\mathbf{z}_i^k$, generated by the GNN, are subsequently fed into two separate prediction heads: f_{atom}, a multi-layer perceptron (MLP) for predicting masked atomic features $\mathbf{h}_i^k$, and f_{pos}, another MLP for predicting masked 3D coordinates R_i^k; and the predictions are $\hat{\mathbf{h}}_i^k = f_{\text{atom}}(\mathbf{z}_i^k)$ and $\hat{R}_i^k = f_{\text{pos}}(\mathbf{z}_i^k)$. For a batch of 3D molecular graphs, the total loss jointly predicts atomic features and spatial coordinates. Let $\mathcal{V}_m^{\text{batch}} = \bigcup_{k=1}^B \mathcal{V}_m^k$ be the set of all masked nodes, and note that CE_Σ denotes the aggregated cross-entropy loss computed over individual masked categorical features. The reconstruction loss for graph $\mathcal{G}^k$ is computed as:

$$\mathcal{L}_{\text{3Dmask}}^k = \sum_{i \in \mathcal{V}_m^k} \frac{\text{CE}_\Sigma\big(\hat{\mathbf{h}}_i^k, \mathbf{h}_i^k\big) + \lambda\, \text{MSE}\big(\hat{R}_i^k, R_i^k\big)}{|\mathcal{V}_m^{\text{batch}}|}. \quad (1)$$

For 2D molecular graphs, which lack spatial information, the masking strategy focuses solely on node and edge feature masking. With the same masking probability p, each node feature $\mathbf{h}_i^k$ and edge feature $\mathbf{e}_{uv}^k$ is independently masked and replaced with zero vectors, with $\mathcal{E}_m^k$ denote the masked edges set. For feature reconstruction, the masked node embeddings $\mathbf{z}_i^k$ are processed by f_{atom} and f_{edge}, another MLP that predicts edge features $\mathbf{e}_{uv}^k$ expressed as $\hat{\mathbf{e}}_{uv}^k = f_{\text{edge}}(\mathbf{z}_u^k, \mathbf{z}_v^k)$. Similarly, the union of masked edges across the batch is $\mathcal{E}_m^{\text{batch}} = \bigcup_{k=1}^B \mathcal{E}_m^k$, and the reconstruction loss for 2D molecular graph $\mathcal{G}^k$ is formulated as:

$$\mathcal{L}_{\text{2Dmask}}^k = \frac{\sum_{i \in \mathcal{V}_m^k} \text{CE}_\Sigma\big(\hat{\mathbf{h}}_i^k, \mathbf{h}_i^k\big)}{|\mathcal{V}_m^{\text{batch}}|} + \frac{\sum_{(u,v) \in \mathcal{E}_m^k} \text{CE}_\Sigma\big(\hat{\mathbf{e}}_{uv}^k, \mathbf{e}_{uv}^k\big)}{|\mathcal{E}_m^{\text{batch}}|}. \quad (2)$$

To generalize across both 2D and 3D molecular graphs, the total masked attribute prediction loss is defined as $\mathcal{L}_{\text{mask}} = \mathcal{L}_{\text{3Dmask}}$, if the input is 3D graph, and $\mathcal{L}_{\text{2Dmask}}$, otherwise.

Molecular descriptor and fingerprint prediction

Molecular descriptors, which quantify global molecular properties, can serve as valuable targets for supervised pretraining tasks. To this end, we calculate six key descriptors using RDKit [Landrum *et al.*, 2024] that are critical for drug discovery: molecular LogP (MolLogP), molecular weight (MolWt), topological polar surface area (TPSA), number of rotatable bonds (NumRotatableBonds), quantitative estimate of drug-likeness (QED), and synthetic accessibility (SA). Furthermore, the molecular fingerprints encapsulate both local and

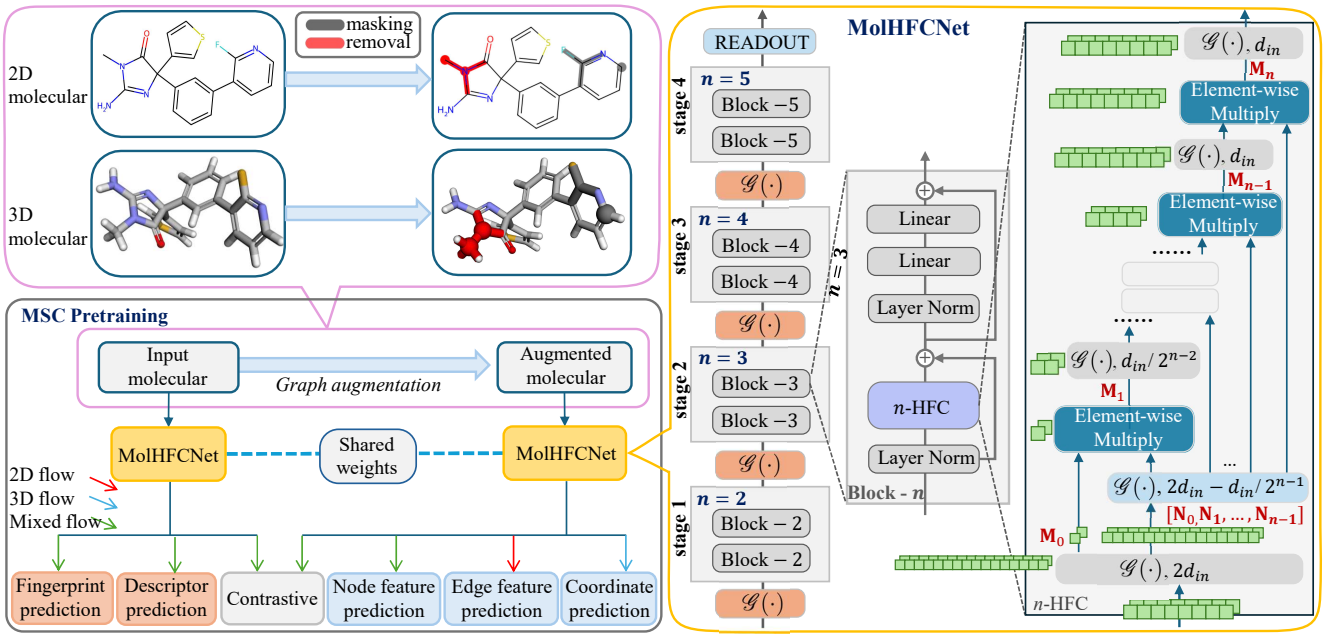

Figure 1: Overview of our pretraining strategy and the architecture of MolHFCNet. The pretraining framework combines predictive learning and contrastive learning, starting with graph augmentations such as subgraph removal and node/edge feature masking. Reconstruction tasks refine embeddings by predicting masked features, while contrastive loss ensures consistency across augmented graph views. Molecular descriptors and fingerprints from RDKit serve as labels for supervised training. The MolHFCNet architecture follows a hierarchical framework for molecular property and interaction prediction, utilizing multiple hierarchical blocks.

global chemical properties such as structural and physicochemical characteristics [Orosz *et al.*, 2022], provide additional complementary information to guide the pretraining process. To predict molecular descriptors and fingerprints, we first generate a graph-level embedding $\mathbf{e}^k$ for each molecular graph $\mathcal{G}^k$ using a readout function. The readout function aggregates node embeddings $\mathbf{z}_i^k$ produced by the GNN into a fixed-size graph representation as $\mathbf{e}^k = \text{READOUT}(\{\mathbf{z}_i^k \mid v_i \in \mathcal{V}^k\})$. This graph-level embedding is then passed through two separate MLPs as $\hat{\mathbf{y}}_d^k = f_{\text{desc}}(\mathbf{e}^k)$ and $\hat{\mathbf{y}}_f^k = f_{\text{fingerprint}}(\mathbf{e}^k)$, where f_{desc} predicts molecular descriptors, and $f_{\text{fingerprint}}$ predicts molecular fingerprints. Let N_d denote the number of descriptors for each graph, with $\{\mathbf{y}_d^k\}$ as ground-truth descriptors and $\{\hat{\mathbf{y}}_d^k\}$ as predictions. The descriptor prediction loss is

$$\mathcal{L}_{\text{desc}} = \frac{1}{B}\sum_{k=1}^{B}\frac{1}{N_d}\sum_{i=1}^{N_d}\text{MSE}\big(\mathbf{y}_{d,i}^k, \hat{\mathbf{y}}_{d,i}^k\big). \tag{3}$$

The fingerprint prediction is considered binary classification tasks. Let N_f denote the fingerprint vector length, and $\{\mathbf{y}_f^k\}$ be the ground-truth fingerprints and $\{\hat{\mathbf{y}}_f^k\}$ be predictions. The binary cross-entropy loss for the batch is

$$\mathcal{L}_{\text{fingerprint}} = \frac{1}{B}\sum_{k=1}^{B}\frac{1}{N_f}\sum_{i=1}^{N_f}\text{CE}\big(\hat{\mathbf{y}}_{f,i}^k, \mathbf{y}_{f,i}^k\big). \tag{4}$$

Contrastive learning

Contrastive learning has emerged as a powerful framework for leveraging the growing abundance of unlabeled datasets, achieving significant success in domains such as computer vision [Chen *et al.*, 2020]. Recently, this paradigm has gained attraction in molecular graph representation learning. The core principle of contrastive learning is to encourage an anchor sample to be closer to its positive samples in the embedding space while pushing it further away from negative samples. In this study, we generate positive samples for molecular graphs by combining attribute masking and subgraph removal techniques [Wang *et al.*, 2022]. Specifically, a positive sample is a pair of an original molecular graph and an augmented graph, resulting in two views: the original graph embedding $\mathbf{e}^k$ and the augmented graph embedding $\mathbf{e}_{\text{aug}}^k$. These embeddings are passed through a shared MLP projection head f_{aug} to generate normalized embeddings as $\hat{\mathbf{e}}^k = f_{\text{aug}}(\mathbf{e}^k)$ and $\hat{\mathbf{e}}_{\text{aug}}^k = f_{\text{aug}}(\mathbf{e}_{\text{aug}}^k)$.

For each batch containing B samples, there are B positive pairs $(\hat{\mathbf{e}}^k, \hat{\mathbf{e}}_{\text{aug}}^k)$ and $B(B-1)$ negative pairs. We then compute the cosine similarity $\text{sim}(\hat{\mathbf{e}}^k, \hat{\mathbf{e}}_{\text{aug}}^k)$ between the original and augmented embeddings. The contrastive loss for original and augmented graph views are defined as:

$$\mathcal{L}_{\text{org}} = -\frac{1}{B}\sum_{k=1}^{B}\log\frac{e^{\text{sim}(\hat{\mathbf{e}}^k, \hat{\mathbf{e}}_{\text{aug}}^k)/\tau}}{\sum_{j=1}^{B} e^{\text{sim}(\hat{\mathbf{e}}^k, \hat{\mathbf{e}}_{\text{aug}}^j)/\tau}}, \tag{5}$$

$$\mathcal{L}_{\text{aug}} = -\frac{1}{B}\sum_{k=1}^{B}\log\frac{e^{\text{sim}(\hat{\mathbf{e}}_{\text{aug}}^k, \hat{\mathbf{e}}^k)/\tau}}{\sum_{j=1}^{B} e^{\text{sim}(\hat{\mathbf{e}}_{\text{aug}}^k, \hat{\mathbf{e}}^j)/\tau}}, \tag{6}$$

where τ is a temperature hyperparameter that controls the sharpness of the similarity distribution. The overall contrastive loss for the batch is the average of the losses under two views as $\mathcal{L}_{\text{contrast}} = \frac{1}{2}(\mathcal{L}_{\text{org}} + \mathcal{L}_{\text{aug}})$.

We utilize a composite loss function combining the objectives for descriptor and fingerprint prediction, node, coordinate and edge recovery, and contrastive learning, which is computed as:

$$\mathcal{L}_{\text{total}} = \mu_1\mathcal{L}_{\text{mask}} + \mu_2\mathcal{L}_{\text{desc}} + \mu_3\mathcal{L}_{\text{fingerprint}} + \mu_4\mathcal{L}_{\text{contrast}}, \quad (7)$$

where $\mu_i \in (0, 1)$, $i = \overline{1,4}$ are weight hyperparameters controlling the relative contributions of each loss component.

2.2 MolHFCNet Architecture

The MolHFCNet architecture is a hierarchical and modular GNN designed to process graph-structured data efficiently by leveraging flexible and expressive GNN layers, denoted generically as $\mathscr{G}(\cdot)$. This modular design allows MolHFCNet to adapt to various graph learning tasks by using different graph layer types, making the architecture highly versatile and flexible. At its core, MolHFCNet employs a multi-order graph convolution mechanism through the n-HFC module, which combines information from multiplying hierarchical feature spaces. This enables the model to aggregate node and edge information dynamically, capturing both local and global graph structures. The architecture has 4 stages, each comprising 2 block modules. Each block integrates the n-HFC, residual connections, and a feed-forward network to ensure stability and expressive representation learning. Figure 1 illustrates the architecture of our proposed MolHFCNet.

The n-HFC module

The n-HFC module is a generalized graph convolution operator designed to efficiently aggregate multi-scale hierarchical information from both node and edge features, enabling flexible modeling of high-order spatial interactions. The order of interactions within the module, denoted by n, represents the depth of hierarchical processing. Inspired by the success of hierarchical convolutional layers in HorNet [Rao *et al.*, 2022] for feature extraction, this module adapts those principles to molecular graph data. The n-HFC module leverages multi-scale message passing to extract local and global structural patterns. Lower layers focus on capturing local neighborhood and atom-level features, while higher layers progressively expand the receptive field to integrate substructural and molecular-level information. This design enables the model to learn fine-grained chemical interactions alongside overarching molecular properties, resulting in richer and more effective graph representations.

Initially, the n-HFC module uses a GNN layer $\mathscr{G}(\cdot)$ to perform feature expansion, transforming the input feature matrix $\mathbf{H} \in \mathbb{R}^{|\mathcal{V}|\times d_{\text{in}}}$ into an expanded feature representation $\mathbf{M} = \mathscr{G}(\mathbf{H}) \in \mathbb{R}^{|\mathcal{V}|\times(2d_{\text{in}})}$. This step expands the feature space, ensuring sufficient capacity to capture local and global patterns simultaneously. The expanded features $\mathbf{M}$ are then decomposed into multiple hierarchical components, including a base feature matrix $\mathbf{M}_0$ and increasing-order features $\{\mathbf{N}_k\}_{k=0}^{n-1}$, which are processed recursively to model interactions at increasing spatial orders, expressed as

$$\mathbf{M} = [\mathbf{M}_0, \mathbf{N}_0, \mathbf{N}_1, \ldots, \mathbf{N}_{n-1}], \quad (8)$$

where $\mathbf{N}_i$ has size of $|\mathcal{V}| \times d_i$ ($\mathbf{M}_0$, $\mathbf{N}_0$ have same size) and

$$d_0 + \sum_{k=0}^{n-1} d_k = 2d_{\text{in}}. \quad (9)$$

Lower-order features, such as $\mathbf{M}_0$, focus on atom-level interactions and local connectivity, while higher-order features expand the receptive field to capture substructural and molecular-level dependencies. This decomposition mirrors the hierarchical principles observed in multi-scale feature extraction from computer vision, enabling the model to adapt effectively to molecular graphs. The n-HFC module continues to process these components by employing recursive gated convolutions, where each step applies $\mathscr{G}(\cdot)$ to expand and transform the features of each order while scaling their contributions dynamically. The recursive updation for the k-th order features, denoted by $\mathbf{M}_{k+1}$, is given by:

$$\mathbf{M}_{k+1} = \frac{\mathscr{G}(\mathbf{N}_k) \odot \mathrm{g}_k(\mathbf{M}_k)}{\xi}, \text{ for } k = 0, 1, \ldots, n-1, \quad (10)$$

where ξ is a scaling factor to normalize the interactions and maintain numerical stability, and $\odot$ denotes the element-wise multiplication. The gating function $\mathrm{g}_k(\cdot)$ ensures compatibility between hierarchical orders by matching the dimensions of features and is defined as the identity mapping iff. $k = 0$, and a trainable $\mathscr{G}(\cdot)$ layer mapping from the feature spaces of $\mathbf{M}_{k-1}$ to $\mathbf{M}_k$ otherwise.

At each step, the receptive field expands, enabling the module to progressively model higher-order spatial interactions. This hierarchical progression ensures that the model captures not only fine-grained chemical interactions at the atom and local neighborhood levels but also global molecular patterns at higher levels of abstraction. After completing the recursion, the output from the final step, $\mathbf{M}_n$, is passed through a projection layer $\mathscr{G}(\cdot)$ to generate the final output of the n-HFC module. This projection consolidates information across all hierarchical orders into a unified representation. Notably, instead of applying different $\mathscr{G}(\cdot)$ at each recursive step, a single $\mathscr{G}(\cdot)$ operation can be performed on the concatenated features $\{\mathbf{N}_k\}_{k=0}^{n-1}$, simplifying the implementation and improving computational efficiency. This efficiency is crucial for scaling to large molecular graphs while maintaining high-order interactions. To further balance expressiveness and computational complexity, the channel dimensions in each hierarchical order are scaled as:

$$d_k = \frac{d_{\text{in}}}{2^{n-k-1}}, \quad \text{for } k = 0, 1, \ldots, n-1. \quad (11)$$

This ensures that lower layers focus on capturing detailed local patterns while higher layers focus on broader molecular structures. By combining hierarchical feature aggregation, efficient recursive computation, and multi-scale message passing, the n-HFC module effectively learns expressive representations of molecular graphs, capturing both local chemical interactions and global structural properties.

The block module

The MolHFCNet architecture comprises four stages, with each stage consisting of two block modules. The block module integrates several key components to enable efficient feature propagation and transformation. Initially, it applies LayerNorm for normalizing input features, followed by a n-HFC module that performs multi-scale message passing. The architecture then proceeds with another layer normalization and a series of linear layers. The block also employs a residual connection to enhance robustness and mitigate overfitting.

Model architectures

We adopt the meta-architecture of HorNet [Rao *et al.*, 2022] as the foundation for constructing MolHFCNet, wherein the fundamental block integrates n-HFC and a feed-forward network. To tailor the model for molecular property and interaction prediction, we simplify the architecture by reducing the number of blocks in each stage to [2, 2, 2, 2]. Additionally, we adapt the base number of hidden dimensions d to construct graph models of varying sizes, setting the number of hidden dimensions across the four stages to $[d, 2d, 4d, 8d]$, in alignment with standard practices for hierarchical graph representation learning.

n-HFC module computational complexity

Theorem 1. *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph with N nodes and $|\mathcal{E}|$ edges. If the cost of full-dimensional $\mathscr{G}(\cdot)\Big|_{d_{\text{in}} \to d_{\text{in}}}$ layer is*

$$\mathcal{C}_{\mathscr{G}} = \mathcal{O}\Big(N\, d_{\text{in}}\, d_{\text{out}} + |\mathcal{E}|\, p(d_{\text{in}}, d_{\text{out}}, d_{\text{e}})\Big), \tag{12}$$

then with the hierarchical scaling $d_k = d_{\text{in}}/2^{n-k-1}$, the total computational complexity of the n-HFC module is

$$\mathcal{C}_{n\text{-HFC}} = \mathcal{O}\Big(\mathcal{C}_{\mathscr{G}}\Big), \tag{13}$$

i.e. approximately the same as that of a single $\mathscr{G}(\cdot)$ layer.

Proof. By analyzing the initial expansion, recursive gated convolutions with scaled feature dimensions, and final projection, the total cost sums to the complexity of a standard $\mathcal{G}(\cdot)$ layer. Detailed proof is omitted due to space limitation. □

The n-HFC module provides substantial advantages in computational efficiency and scalability, making it well-suited for deep graph networks. As established in Theorem 1, its total computational complexity remains comparable to that of a single $\mathscr{G}$ layer, despite performing hierarchical feature expansion across multiple levels. This efficiency is achieved through a structured scaling of feature dimensions, where each successive layer expands the feature space by a factor of 2, effectively balancing expressive power and computational cost. By maintaining a complexity of $\mathcal{O}(\mathcal{C}_{\mathscr{G}})$, the n-HFC module enables deeper, multi-hop feature extraction compared to the single-hop processing of a standard $\mathscr{G}$ layer, without introducing significant computational overhead. Moreover, its adaptable architecture seamlessly integrates with various graph neural network backbones, making it a robust and scalable solution for molecular property prediction and other graph-based tasks. Thus, the n-HFC module offers a principled and efficient approach to constructing deep hierarchical graph networks that effectively balance computational complexity with rich, multi-scale feature learning.

3 Experiments

3.1 Backbone Layers

Our proposed model is designed to be modular and flexible, allowing the use of different GNN layers as the fundamental building blocks. Instead of being restricted to a specific architecture, our method supports a variety of GNN layers, including but not limited to Graph Attention Networks (GATs, in [Veličković *et al.*, 2018]), Graph Isomorphism Networks (GINs, in [Xu *et al.*, 2019]), Graph Convolutional Networks (GCNs, in [Kipf and Welling, 2017]), Graph Transformers Networks (GTNs, in [Shi *et al.*, 2020]), Spatial Graph Convolutional Networks (SGCNs, in [Danel *et al.*, 2020]) and Continuous-Filter Convolutions (CFConv) from SchNet [2017]. These layers serve as interchangeable components within our model, enabling it to adapt to both 2D molecular graphs and 3D molecular conformations.

Other GNN Layers. We highlight that our proposed framework is flexible to $\mathscr{G}(\cdot)$ and can be extended to incorporate more advanced architectures, such as DimeNet [2020], PaiNN [2021] and GemNet [2021] . However, certain architectures substantially increase the number of parameters and demand significant computational resources, making them less practical in specific settings. A balance between model performance, computational efficiency, and memory constraints therefore guides the selection of layers.

3.2 Implementation Settings

We pretrained the MolHFCNet model on a dataset of 10 million SMILES strings from the PubChem database [2020], which were provided by [Chithrananda *et al.*, 2020] for 5 epochs. For molecular data in SMILES format, we employ RDKit [Landrum *et al.*, 2024] to generate simulated 3D coordinates, along with molecular descriptors and fingerprints. To optimize performance, we carefully tuned the hyperparameters during training. Batch sizes were selected from {8, 16, 32}, and learning rates were chosen from $\{1\text{e}{-4}, 5\text{e}{-4}, 1\text{e}{-3}\}$. A cosine annealing schedule with a warmup phase was employed to gradually decrease the learning rate. The model was trained for {60, 80, 100} epochs, with early stopping criteria set at {12, 16, 20} epochs to prevent overfitting. All training experiments were conducted on a single Tesla V100 GPU with 32 GB of memory.

3.3 Molecular Property Prediction Tasks

Datasets

To evaluate the effectiveness of our proposed framework, we performed extensive experiments on nine benchmark datasets from MoleculeNet [Wu *et al.*, 2018] for molecular properties prediction including ESOL, FreeSolv, Lipophilicity, BACE, BBBP, HIV, ClinTox, SIDER, and Tox21. In this work, we follow the data curation and splitting [Nguyen *et al.*, 2024].

Baselines

We evaluated the performance of our proposed method by comparing it against several baseline models, including 2D-GNNs such as GAT and GCN. Furthermore, we benchmarked against advanced 2D-GNN-based molecular property prediction models, such as AttentiveFP [2019], GMT [2021], TrimNet [2021], D-MPNN [2019], HiGNN [2022], and ResGAT [2024], which do not utilize pretraining. For pretraining-based approaches, we included models like HiMol [2023] and MolCLR [2022]. Additionally, we assessed Transformer-based models using SMILES representations, including ChemBERTa [2020]. Other models using 3D

		Regression Dataset				Classification Dataset						
Sampling	Model	FreeSolv	ESOL	Lipophilicity	Avg.	BACE	BBBP	HIV	SIDER	Tox21	Clintox	Avg.
Random	**Previous studies**											
	AttentiveFP	3.79 (0.64)	1.58 (0.13)	1.14 (0.03)	2.17	89.25 (2.2)	88.73 (2.2)	79.54 (1.9)	64.51 (2.5)	85.44 (1.0)	89.46 (5.0)	82.82
	D-MPNN	1.14 (0.23)	0.74 (0.10)	0.62 (0.02)	0.83	89.24 (2.6)	89.62 (2.6)	80.43 (1.4)	63.17 (0.9)	85.38 (1.2)	88.18 (4.8)	82.67
	HiGNN	5.82 (0.94)	3.80 (0.29)	1.68 (0.03)	3.76	89.41 (2.6)	90.78 (2.3)	78.28 (1.9)	62.51 (0.4)	85.73 (1.9)	91.34 (9.0)	83.01
	ResGAT	1.40 (0.25)	0.77 (0.07)	0.67 (0.01)	0.95	83.47 (8.7)	86.49 (2.6)	77.54 (0.7)	61.66 (3.4)	83.93 (3.5)	88.75 (4.2)	80.31
	ChemBERTa	1.35 (0.22)	0.76 (0.08)	0.73 (0.04)	0.95	86.24 (0.5)	86.84 (2.2)	72.49 (1.7)	62.61 (3.2)	83.47 (2.1)	86.28 (5.3)	79.65
	HiMol	1.36 (0.26)	0.68 (0.09)	0.66 (0.03)	0.90	88.38 (3.1)	87.96 (2.8)	76.00 (3.1)	59.72 (0.9)	85.38 (1.6)	78.44 (8.5)	79.31
	MolCLR	1.80 (0.58)	0.93 (0.11)	0.66 (0.04)	1.13	86.20 (4.1)	91.25 (3.4)	77.33 (2.9)	61.29 (1.2)	81.76 (1.4)	88.89 (2.9)	81.12
	SGCN	2.25 (0.38)	1.46 (0.13)	1.09 (0.03)	1.60	71.09 (8.2)	72.34 (5.0)	70.61 (2.1)	56.39 (1.9)	–	60.61 (10.8)	–
	SchNet	1.26 (0.26)	0.64 (0.09)	0.62 (0.03)	0.84	88.38 (3.1)	87.96 (2.8)	76.00 (3.1)	59.72 (0.9)	85.38 (1.6)	78.44 (8.5)	79.31
	Uni-Mol	0.86 (0.15)	0.60 (0.06)	**0.54 (0.01)**	**0.67**	87.22 (4.3)	90.06 (2.1)	78.40 (1.6)	61.90 (3.6)	86.10 (1.8)	83.5 (1.2)	81.20
	Our models											
	MolHFCNet-GAT	1.03 (0.18)	0.79 (0.06)	0.64 (0.02)	0.82	**90.37 (3.5)**	91.44 (2.8)	80.27 (0.5)	65.3 (1.9)	85.97 (2.7)	93.89 (4.2)	84.54
	MolHFCNet-GCN	1.06 (0.18)	0.76 (0.04)	0.65 (0.02)	0.82	89.15 (2.7)	90.66 (3.1)	**81.30 (1.5)**	66.62 (2.9)	85.82 (2.2)	91.40 (6.2)	84.16
	MolHFCNet-GIN	0.92 (0.08)	0.69 (0.02)	0.63 (0.02)	0.75	90.27 (1.7)	91.98 (1.8)	80.46 (2.0)	65.80 (2.4)	86.43 (3.0)	94.12 (4.7)	84.84
	MolHFCNet-GTN	1.13 (0.34)	0.67 (0.06)	0.66 (0.01)	0.82	90.26 (1.3)	91.25 (1.6)	81.27 (2.2)	**66.76 (1.6)**	**86.55 (1.7)**	95.02 (3.3)	**85.18**
	MolHFCNet-SGCN	0.98 (0.21)	0.61 (0.07)	0.62 (0.01)	0.74	90.36 (2.0)	91.60 (1.5)	80.56 (1.8)	64.60 (2.8)	–	93.91 (3.5)	–
	MolHFCNet-CFC	**0.83 (0.19)**	**0.58 (0.06)**	0.60 (0.03)	**0.67**	89.82 (1.8)	**92.38 (1.2)**	80.21 (1.1)	64.13 (1.8)	85.66 (2.5)	**95.37 (2.1)**	84.60
Scaffold	**Previous studies**											
	AttentiveFP	4.99 (0.47)	1.74 (0.36)	1.11 (0.07)	2.61	81.74 (5.1)	87.86 (5.6)	76.92 (7.4)	58.77 (3.6)	82.23 (2.2)	75.65 (16.0)	77.19
	D-MPNN	1.88 (0.47)	0.91 (0.13)	0.64 (0.03)	1.14	81.60 (4.5)	90.28 (3.8)	77.92 (7.4)	58.81 (7.0)	81.54 (2.6)	80.19 (13.1)	78.39
	HiGNN	47.11 (1.25)	4.51 (0.94)	1.68 (0.05)	4.44	84.54 (2.3)	86.12 (1.3)	77.64 (2.2)	60.28 (3.8)	81.57 (1.8)	79.54 (16.7)	78.28
	ResGAT	11.89 (0.44)	1.09 (0.16)	0.71 (0.03)	1.23	75.49 (7.9)	87.11 (5.0)	73.26 (5.1)	61.40 (2.9)	79.83 (3.5)	81.09 (3.3)	76.36
	ChemBERTa	2.96 (0.41)	1.13 (0.19)	0.80 (0.03)	1.63	81.41 (4.0)	88.43 (4.7)	69.57 (4.2)	60.80 (1.9)	78.48 (1.6)	83.90 (6.2)	77.10
	HiMol	2.93 (0.28)	0.87 (0.05)	0.70 (0.04)	1.50	82.44 (4.2)	88.86 (4.8)	75.53 (6.1)	57.78 (4.3)	80.81 (1.7)	66.19 (5.6)	75.27
	MolCLR	2.47 (0.53)	1.28 (0.08)	0.65 (0.05)	1.47	82.84 (3.4)	87.66 (4.6)	73.71 (6.4)	58.13 (1.5)	78.43 (1.6)	85.74 (3.6)	77.75
	SGCN	2.86 (0.49)	1.82 (0.47)	1.08 (0.06)	1.92	72.37 (5.3)	75.76 (5.3)	71.24 (4.8)	58.82 (2.7)	–	63.17 (4.4)	–
	SchNet	2.73 (0.28)	0.87 (0.05)	0.66 (0.04)	1.42	83.44 (4.2)	89.86 (4.8)	76.63 (6.1)	59.78 (4.3)	82.81 (1.7)	68.19 (5.6)	76.70
	Uni-Mol	1.73 (0.37)	**0.78 (0.08)**	**0.57 (0.04)**	1.03	83.57 (3.7)	87.86 (3.7)	76.58 (5.4)	61.23 (1.6)	81.40 (1.5)	80.73 (8.8)	78.56
	Our models											
	MolHFCNet-GAT	2.02 (0.37)	1.09 (0.11)	0.69 (0.04)	1.27	84.88 (3.3)	88.62 (4.4)	77.75 (5.0)	63.91 (3.2)	80.49 (2.7)	89.84 (1.9)	80.92
	MolHFCNet-GCN	2.77 (0.88)	1.04 (0.16)	0.69 (0.04)	1.50	85.11 (2.3)	89.49 (3.1)	76.16 (4.3)	63.09 (2.1)	81.30 (2.1)	87.07 (4.1)	80.37
	MolHFCNet-GIN	1.86 (0.72)	0.95 (0.05)	0.66 (0.03)	1.16	85.07 (3.1)	89.80 (2.6)	76.78 (6.4)	**63.96 (2.7)**	82.17 (2.0)	89.54 (5.8)	81.22
	MolHFCNet-GTN	2.14 (0.55)	0.94 (0.08)	0.70 (0.05)	1.26	84.15 (3.3)	90.56 (3.0)	77.49 (5.0)	62.86 (2.6)	**82.25 (3.3)**	89.28 (8.5)	81.10
	MolHFCNet-SGCN	1.79 (0.59)	0.87 (0.11)	0.66 (0.02)	1.16	84.58 (3.5)	90.44 (4.5)	77.45 (4.2)	62.31 (2.2)	–	90.38 (5.1)	–
	MolHFCNet-CFC	**1.63 (0.41)**	0.83 (0.10)	0.61 (0.03)	**1.02**	**85.29 (2.3)**	**90.67 (3.2)**	**78.50 (5.0)**	60.71 (2.3)	81.37 (4.2)	**92.95 (2.2)**	**81.58**

Table 1: The average RMSE and ROC-AUC results (with standard deviation) on MoleculeNet's regression and classification test sets. The best results are highlighted in bold, with the second-best results underlined.

Graph information, including SchNet [2017], SGCN [2020] and Uni-Mol [2023], were also included in our experiments.

Performance evaluation

The results in Table 1 demonstrates significant improvements of our proposed MolHFCNets over previous studies, particularly in both regression and classification datasets. In the regression tasks under random splitting, MolHFCNet-CFC achieves the best average RMSE (0.67), outperforming all baseline models, including SchNet (0.84) and competitive with Uni-Mol (0.67) with 44M parameters, while securing the lowest RMSE in ESOL (0.58) and FreeSolv (0.83). Similarly, under scaffold splitting, MolHFCNet-CFC maintains its superiority with the best overall RMSE (1.02), significantly surpassing previous best models such as Uni-Mol and D-MPNN.

In classification tasks, MolHFCNet consistently delivers top-tier results, particularly in key datasets. Under random splitting, MolHFCNet-GTN achieves the highest average ROC-AUC, while MolHFCNet-CFC attains the best ROC-AUC in BBBP (92.38) and Clintox (95.37) and demonstrates competitive performance across other datasets, resulting in a strong overall average of 84.6. Under scaffold splitting, MolHFCNet-CFC remains the top performer, achieving the best overall ROC-AUC (81.58), excelling in BBBP (90.67), HIV (78.5), and Clintox (92.95), clearly outperforming previous models such as D-MPNN (78.39) and Uni-Mol (78.56). These results highlight the robustness and effectiveness of MolHFCNet across diverse datasets and tasks.

Model	LBA (RMSE) ↓	LBA (R_P) ↑	LBA (R_S) ↑	LEP (ROC) ↑	LEP (PR) ↑
GeoSSL-RR	1.515 ± 0.07	0.545 ± 0.03	0.539 ± 0.03	0.654 ± 0.05	0.518 ± 0.06
GeoSSL-InfoNCE	1.564 ± 0.05	0.508 ± 0.03	0.497 ± 0.05	0.693 ± 0.06	0.571 ± 0.08
GeoSSL-EBM-NCE	1.499 ± 0.06	0.547 ± 0.03	0.534 ± 0.03	0.691 ± 0.05	0.603 ± 0.07
GeoSSL-DDM	1.451 ± 0.03	$\mathbf{0.577 \pm 0.02}$	$\mathbf{0.572 \pm 0.01}$	0.776 ± 0.03	0.694 ± 0.06
MolHFCNet-SGCN	1.276 ± 0.01	0.457 ± 0.01	0.430 ± 0.01	0.627 ± 0.08	0.503 ± 0.06
MolHFCNet-CFC	$\mathbf{1.238 \pm 0.03}$	0.508 ± 0.02	0.490 ± 0.02	$\mathbf{0.801 \pm 0.01}$	$\mathbf{0.756 \pm 0.02}$

Table 2: Comparison of Binding Affinity Prediction results. We report Root Mean Squared Error (RMSE), Pearson correlation (R_P), and Spearman correlation (R_S) for LBA, while LEP is evaluated using ROC-AUC and PR-AUC.

3.4 Binding Affinity Prediction Tasks

Datasets

In this study, we follow the binding affinity prediction tasks as described in [Liu *et al.*, 2022a] to evaluate molecular interactions, crucial for molecular docking and drug-target studies. Using the Atom3D dataset [Townshend *et al.*, 2020], we assess our model's performance on two key tasks: Ligand Binding Affinity (LBA), which predicts ligand-protein interaction strength, and Ligand Efficacy Prediction (LEP), which determines whether a ligand exhibits a stronger binding affinity toward one protein pocket compared to another.

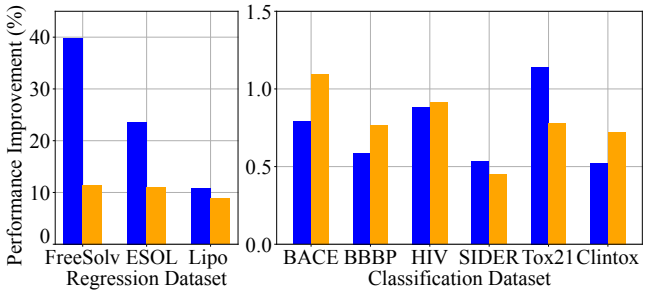

Figure 2: Average performance improvement of pretrained MolHFCNet models on molecular property prediction tasks.

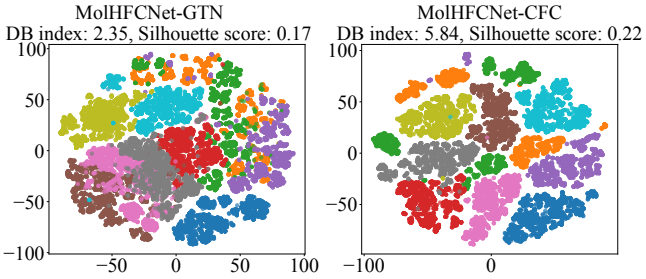

Figure 3: Visualization of t-SNE representations learned by MolHFCNet-GTN and MolHFCNet-CFC in different scaffolds.

Baselines

To evaluate the effectiveness of our approach, we compare it against four state-of-the-art models: GeoSSL-RR [2022b], GeoSSL-InfoNCE and GeoSSL-EBM-NCE and GeoSSL-DDM [2022a], all evaluated using the PaiNN backbone architecture. For a fair comparison, we adopt the same experimental settings and data splitting strategy as [Liu *et al.*, 2022a]. Baseline model results are obtained from [Liu *et al.*, 2022a].

Performance evaluation

The results in Table 2 demonstrates their competitive edge, particularly in achieving superior results in key metrics. MolHFCNet-CFC outperforms all GeoSSL-based models in terms of LBA RMSE, achieving the lowest error (1.238 ± 0.03), significantly improving upon GeoSSL-RR (1.515 ± 0.07) and GeoSSL-DDM (1.451 ± 0.03). Although GeoSSL-DDM maintains the highest Pearson (R_P) and Spearman (R_S) correlations, MolHFCNet-CFC delivers competitive results. In LEP evaluation, MolHFCNet-CFC achieves the best performance, setting new benchmarks in both ROC-AUC (0.801 ± 0.01) and PR-AUC (0.756 ± 0.02), outperforming all GeoSSL models, including the previously best GeoSSL-DDM (0.776 ± 0.03 ROC and 0.694 ± 0.06 PR). These results highlight the effectiveness of our approach in affinity prediction, demonstrating that MolHFCNet, especially the CFC variant, is a highly promising method for binding affinity prediction, surpassing existing models in predictive capabilities.

3.5 Impact of Pretraining on Performance

The impact of pretraining on performance, as illustrated in Figure 2, highlights the benefits of pretraining for MolHFCNet models across both regression and classification tasks in molecular property prediction. The performance improvement trends show that pretraining has a more pronounced effect under the random split setting, particularly in regression datasets such as FreeSolv, ESOL, and Lipophilicity, where improvements are significantly higher compared to the scaffold split. For classification tasks, the performance gains from pretraining are more subtle, with improvements generally under 2%. However, MolHFCNet consistently demonstrates a positive impact across all classification datasets, with notable gains in datasets like BBBP and Tox21 under the random split setting. The relatively smaller improvements in classification datasets compared to regression datasets suggest that pretraining contributes more significantly to learning continuous molecular property relationships than binary classification tasks. Overall, this analysis confirms that pretraining enhances the predictive performance of MolHFCNet, particularly under the random split setting, reinforcing its effectiveness in molecular property prediction tasks.

3.6 Visualization of Model Representation

To evaluate whether the pretrained representations effectively capture scaffold information—reflecting the core structural frameworks of bioactive compounds [Hu *et al.*, 2016]—we applied t-SNE [Van der Maaten and Hinton, 2008] to visualize embeddings from the two best-performing models: MolHFCNet-GTN (left) and MolHFCNet-CFC (right), as shown in Figure 3. Following [Li *et al.*, 2020; Zhu *et al.*, 2023], we selected the ten most common scaffolds from the ZINC15 database [Sterling and Irwin, 2015] and randomly sampled 2,000 molecules per scaffold, yielding 20,000 compounds for analysis. MolHFCNet-GTN achieved a lower DB index [Davies and Bouldin, 1979] (2.35 vs. 5.84), indicating more compact clustering, whereas MolHFCNet-CFC obtained a higher silhouette score [Shahapure and Nicholas, 2020] (0.22 vs. 0.17), suggesting better-defined cluster separation. Notably, while CFC representations exhibited clearer clustering, the dispersed distribution of cluster 3 contributed to a higher DB index, whereas GTN representations displayed greater overlap between clusters.

4 Conclusions

In this work, we introduced MolHFCNet, a versatile GNN architecture designed for molecular property and interaction prediction, capable of leveraging either 2D or 3D molecular graph representations. At its core, the n-HFC module enables multi-hop feature extraction while maintaining computational efficiency comparable to a single standard graph layer. The model integrates a hierarchical multi-scale representation strategy and a novel graph pretraining framework, combining predictive and contrastive learning to enhance molecular embeddings. Experimental results demonstrate MolHFCNet's superiority over baseline methods across molecular property and binding affinity prediction tasks. Future work includes enhancing model robustness with advanced 3D graph layers, integrating both 2D and 3D representations, and expanding applications to molecular generation and optimization tasks.

Acknowledgments

This work was supported in part by the Endeavour Fund from the New Zealand Ministry of Business, Innovation and Employment (MBIE) under contract RTVU2301.

References

[Alberga *et al.*, 2024] Domenico Alberga, Giuseppe Lamanna, Giovanni Graziano, Pietro Delre, Maria Cristina Lomuscio, et al. DeLA-DrugSelf: Empowering multi-objective de novo design through SELFIES molecular representation. *Computers in Biology and Medicine*, 175:108486, 2024.

[Baek *et al.*, 2021] Jinheon Baek, Minki Kang, and Sung Ju Hwang. Accurate learning of graph representations with graph multiset pooling. *arXiv:2102.11533*, 2021.

[Chen *et al.*, 2020] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning*, pages 1597–1607. PMLR, 2020.

[Chithrananda *et al.*, 2020] Seyone Chithrananda, Gabriel Grand, and Bharath Ramsundar. ChemBERTa: Large-scale self-supervised pretraining for molecular property prediction. *arXiv:2010.09885*, 2020.

[Danel *et al.*, 2020] Tomasz Danel, Przemysław Spurek, Jacek Tabor, Marek Śmieja, Łukasz Struski, Agnieszka Słowik, and Łukasz Maziarka. Spatial graph convolutional networks. In *International Conference on Neural Information Processing*, pages 668–675. Springer, 2020.

[Davies and Bouldin, 1979] David L Davies and Donald W Bouldin. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227, 1979.

[Duan *et al.*, 2022] Keyu Duan, Zirui Liu, Peihao Wang, Wenqing Zheng, Kaixiong Zhou, Tianlong Chen, Xia Hu, and Zhangyang Wang. A comprehensive study on large-scale graph training: Benchmarking and rethinking. *Advances in Neural Information Processing Systems*, 35:5376–5389, 2022.

[Gasteiger *et al.*, 2020] Johannes Gasteiger, Shankari Giri, Johannes T. Margraf, and Stephan Günnemann. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. *arXiv:2011.14115*, 2020.

[Gilmer *et al.*, 2017] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International Conference on Machine Learning*, pages 1263–1272. PMLR, 2017.

[Hu *et al.*, 2016] Ye Hu, Dagmar Stumpfe, and Jürgen Bajorath. Computational exploration of molecular scaffolds in medicinal chemistry. *Journal of Medicinal Chemistry*, 59(9):4062–4076, 2016.

[Hu *et al.*, 2020] Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. Strategies for pre-training graph neural networks. In *International Conference on Learning Representations*, 2020.

[Kim *et al.*, 2020] Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A Shoemaker, Paul A Thiessen, Bo Yu, Leonid Zaslavsky, Jian Zhang, and Evan E Bolton. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Research*, 49(D1):D1388–D1395, 2020.

[Kipf and Welling, 2017] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

[Klicpera *et al.*, 2021] Johannes Klicpera, Florian Becker, and Stephan Günnemann. GemNet: Universal directional graph neural networks for molecules. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, pages 6790–6802, 2021.

[Landrum *et al.*, 2024] Greg Landrum, Paolo Tosco, Brian Kelley, Ricardo Rodriguez, David Cosgrove, et al. RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling, 2024.

[Li *et al.*, 2020] Pengyong Li, Jun Wang, Yixuan Qiao, Hao Chen, Yihuan Yu, Xiaojun Yao, Peng Gao, Guotong Xie, and Sen Song. Learn molecular representations from large-scale unlabeled molecules for drug discovery. *arXiv:2012.11175*, 2020.

[Li *et al.*, 2021] Pengyong Li, Yuquan Li, Chang-Yu Hsieh, Shengyu Zhang, Xianggen Liu, Huanxiang Liu, Sen Song, and Xiaojun Yao. TrimNet: learning molecular representation from triplet messages for biomedicine. *Briefings in Bioinformatics*, 22(4):bbaa266, 2021.

[Liu *et al.*, 2022a] Shengchao Liu, Hongyu Guo, and Jian Tang. Molecular geometry pretraining with SE (3)-invariant denoising distance matching. *arXiv:2206.13602*, 2022.

[Liu *et al.*, 2022b] Shengchao Liu, Hanchen Wang, Weiyang Liu, Joan Lasenby, Hongyu Guo, and Jian Tang. Pre-training molecular graph representation with 3D geometry. In *International Conference on Learning Representations*, 2022.

[Nguyen *et al.*, 2024] Long D. Nguyen, Quang H. Nguyen, Quang H. Trinh, and Binh P. Nguyen. From SMILES to enhanced molecular property prediction: A unified multimodal framework with predicted 3D conformers and contrastive learning techniques. *Journal of Chemical Information and Modeling*, 64(24):9173–9195, 2024.

[Nguyen-Vo *et al.*, 2024] Thanh-Hoang Nguyen-Vo, Trang T. T. Do, and Binh P. Nguyen. ResGAT: Residual graph attention networks for molecular property prediction. *Memetic Computing*, 16:491–503, 2024.

[Nigam *et al.*, 2020] AkshatKumar Nigam, Pascal Friederich, Mario Krenn, and Alan Aspuru-Guzik. Augmenting genetic algorithms with deep neural networks for exploring the chemical space. In *International Conference on Learning Representations*, 2020.

[Orosz *et al.*, 2022] Álmos Orosz, Károly Héberger, and Anita Rácz. Comparison of descriptor-and fingerprint sets in machine learning models for ADME-Tox targets. *Frontiers in Chemistry*, 10:852893, 2022.

[Rao *et al.*, 2022] Yongming Rao, Wenliang Zhao, Yansong Tang, Jie Zhou, Ser Nam Lim, and Jiwen Lu. HorNet: Efficient high-order spatial interactions with recursive gated convolutions. In *Advances in Neural Information Processing Systems*, volume 35, pages 10353–10366, 2022.

[Schütt *et al.*, 2017] Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in Neural Information Processing Systems*, 30, 2017.

[Schütt *et al.*, 2021] Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, pages 9377–9388. PMLR, 2021.

[Shahapure and Nicholas, 2020] Ketan Rajshekhar Shahapure and Charles Nicholas. Cluster quality analysis using silhouette score. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 747–748. IEEE, 2020.

[Shi *et al.*, 2020] Yunsheng Shi, Zhengjie Huang, Shikun Feng, Hui Zhong, Wenjin Wang, and Yu Sun. Masked label prediction: Unified message passing model for semi-supervised classification. *arXiv:2009.03509*, 2020.

[Sterling and Irwin, 2015] Teague Sterling and John J Irwin. ZINC 15–ligand discovery for everyone. *Journal of Chemical Information and Modeling*, 55(11):2324–2337, 2015.

[Thiede *et al.*, 2022] Luca A. Thiede, Mario Krenn, AkshatKumar Nigam, and Alán Aspuru-Guzik. Curiosity in exploring chemical spaces: intrinsic rewards for molecular reinforcement learning. *Machine Learning: Science and Technology*, 3(3):035008, 2022.

[Townshend *et al.*, 2020] Raphael JL Townshend, Martin Vögele, Patricia Suriana, Alexander Derry, Alexander Powers, Yianni Laloudakis, Sidhika Balachandar, Bowen Jing, Brandon Anderson, Stephan Eismann, et al. Atom3D: Tasks on molecules in three dimensions. *arXiv:2012.04035*, 2020.

[Van der Maaten and Hinton, 2008] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(11), 2008.

[Veličković *et al.*, 2018] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.

[Wang *et al.*, 2019] Sheng Wang, Yuzhi Guo, Yuhong Wang, Hongmao Sun, and Junzhou Huang. 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*, BCB '19. ACM, 2019.

[Wang *et al.*, 2022] Yuyang Wang, Jianren Wang, Zhonglin Cao, and Amir Barati Farimani. Molecular contrastive learning of representations via graph neural networks. *Nature Machine Intelligence*, 4(3):279–287, 2022.

[Wu *et al.*, 2018] Zhenqin Wu, Bharath Ramsundar, Evan N. Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S. Pappu, Karl Leswing, and Vijay Pande. MoleculeNet: a benchmark for molecular machine learning. *Chemical Science*, 9(2):513–530, 2018.

[Xiong *et al.*, 2019] Zhaoping Xiong, Dingyan Wang, Xiaohong Liu, Feisheng Zhong, et al. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *Journal of Medicinal Chemistry*, 63(16):8749–8760, 2019.

[Xu *et al.*, 2019] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019.

[Yang *et al.*, 2019] Kevin Yang, Kyle Swanson, Wengong Jin, Connor Coley, Philipp Eiden, et al. Analyzing learned molecular representations for property prediction. *Journal of Chemical Information and Modeling*, 59(8):3370–3388, 2019.

[Yüksel *et al.*, 2023] Atakan Yüksel, Erva Ulusoy, Atabey Ünlü, and Tunca Doğan. SELFormer: molecular representation learning via SELFIES language models. *Machine Learning: Science and Technology*, 4(2):025035, 2023.

[Zang *et al.*, 2023] Xuan Zang, Xianbing Zhao, and Buzhou Tang. Hierarchical molecular graph self-supervised learning for property prediction. *Communications Chemistry*, 6(1):34, 2023.

[Zhou *et al.*, 2023] Gengmo Zhou, Zhifeng Gao, Qiankun Ding, Hang Zheng, Hongteng Xu, Zhewei Wei, Linfeng Zhang, and Guolin Ke. Uni-Mol: A universal 3D molecular representation learning framework. In *International Conference on Learning Representations*, 2023.

[Zhu *et al.*, 2022] Weimin Zhu, Yi Zhang, Duancheng Zhao, Jianrong Xu, and Ling Wang. HiGNN: A hierarchical informative graph neural network for molecular property prediction equipped with feature-wise attention. *Journal of Chemical Information and Modeling*, 63(1):43–55, 2022.

[Zhu *et al.*, 2023] Jinhua Zhu, Yingce Xia, Lijun Wu, Shufang Xie, Wengang Zhou, Tao Qin, Houqiang Li, and Tie-Yan Liu. Dual-view molecular pre-training. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '23. ACM, 2023.