

A hierarchical interaction message net for accurate molecular property prediction

Received: 27 May 2025

Accepted: 23 January 2026

Cite this article as: Hong, H., Wu, X., Sun, H. *et al.* A hierarchical interaction message net for accurate molecular property prediction. *Commun Chem* (2026). <https://doi.org/10.1038/s42004-026-01922-x>

Huiyang Hong, Xinkai Wu, Hongyu Sun, Chaoyang Xie, Qi Wang & Yuquan Li

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

A hierarchical interaction message net for accurate molecular property prediction

Huiyang Hong Xinkai Wu Hongyu Sun Chaoyang Xie Qi Wang* Yuquan Li*

H. Hong, X. Wu, H. Sun, C. Xie, Prof. Q. Wang, Prof. Y. Li.

State Key Laboratory of Public Big Data, College of Computer Science and Technology, Guizhou University, 550025 Guizhou, China

Email Address: sdc.hyhong22@gzu.edu.cn (Huiyang Hong), cse.xkwu22@gzu.edu.cn (Xinkai Wu), cse.hysun22@gzu.edu.cn (Hongyu Sun), gs.cyxie23@gzu.edu.cn (Chaoyang Xie), qiwang@gzu.edu.cn (Qi Wang), yvquan.li@gzu.edu.cn (Yuquan Li)

Keywords: *Molecular property prediction, Graph Neural Networks, Motif interaction*

Discovering molecules with desirable molecular properties, including ADMET profiles, is of great importance in drug discovery. Existing approaches typically employ deep learning models, such as Graph Neural Networks and Transformers, to predict these molecular properties by learning from diverse chemical information. However, these models often lack mechanisms for effective interaction among multi-level features. To address these limitations, we propose a Hierarchical Interaction Message Passing Mechanism, which serves as the foundation of our model, the Hierarchical Interaction Message Net (HimNet). Our method enables interaction-aware representation learning across atomic, motif, and molecular levels via hierarchical attention-guided message passing. This design allows HimNet to effectively balance global and local information, ensuring rich and task-relevant feature extraction for downstream property prediction tasks. We systematically evaluate HimNet on eleven datasets, including eight widely-used MoleculeNet benchmarks and three challenging, high-value datasets for metabolic stability, malaria activity, and liver microsomal clearance, covering a broad range of pharmacologically relevant properties. Extensive experiments demonstrate that HimNet achieves the best or near-best performance in most molecular property prediction tasks. We believe that HimNet offers an accurate and efficient solution for molecular activity and ADMET property prediction, contributing significantly to advanced decision-making in the early stages of drug discovery.

1 Introduction

Accurate prediction of molecular properties remains a crucial task in small-molecule drug discovery [1]. Early assessment of features such as bioactivity, solubility, permeability, and toxicity enables medicinal chemists to efficiently filter out unsuitable compounds from large screening libraries, thereby reducing the need for costly and time-consuming experimental assays [2]. Although *in vitro* and *in vivo* experiments remain the gold standard for property validation, their low throughput and high expense limit their practicality for large-scale screening. Consequently, there is a growing shift toward computational, data-driven strategies that accelerate lead optimization and mitigate downstream risks in drug development [3].

Moreover, graph neural networks (GNNs) have emerged as a leading computational strategy for molecular property prediction[3]. By representing atoms as graph nodes and bonds as edges, GNNs jointly capture local chemical environments and global topological context without reliance on manually engineered descriptors or fixed fingerprints. Seminal architectures, such as graph convolutional networks (GCN)[4], graph attention networks (GAT)[5] and message passing neural networks (MPNN)[6] have consistently outperformed classical machine-learning baselines. Subsequent refinements, including directed MPNN (D-MPNN)[7], contextual MPNN (C-MPNN)[8], Graphormer[9] and fragment-aware GNNs[10] have further enriched representational power by integrating edge features, bond directionality or fragment-level encodings. In addition, hybrid models combining graph encoders with global attention mechanisms[11, 12] have demonstrated broad applicability across diverse ADMET (absorption, distribution, metabolism, excretion, and toxicity) benchmarks, underscoring the maturity and generality of hierarchical GNN approaches in the field.

However, despite this progress, virtually all existing GNN variants remain confined to single-layer information flows—either atomic or fragmental—thus failing to model dynamic, context-dependent interactions among functional groups that critically influence molecular behavior[13, 14]. Recent chemical

studies have shown that such synergistic effects are far from additive: for example, K. Chuang *et al.* introduced the substructure masking explanation (SME) framework[15] to attribute property variations to specific motif combinations, revealing that randomly masking paired substructures produces markedly different attribution patterns than masking each motif in isolation. Similarly, Costa *et al.*[16] demonstrated non-linear enhancement of photodynamic-therapy activity in hydrazone-functionalized corroles only when hydrazone and N-Boc/N-Ts groups co-occur, and Dawood *et al.*[17] found that 5-sulfanyl thiazole derivatives exhibit potent anticancer activity only with particular aromatic substituent pairs. Moreover, Breaking of Retrosynthetically Interesting Chemical Substructures(BRICS) decomposition[18] can efficiently identify chemically meaningful motifs that correlate with target properties, yet current GNN-based methods typically aggregate these motifs via simple pooling or static embeddings, thereby overlooking their cooperative interplay. These observations highlight an urgent need for a model that not only processes hierarchical information but also enables learnable, cross-layer interactions among motifs and functional groups—only then can we fully capture the complex determinants of ADMET properties[13, 14].

To address these challenges, we propose a novel, to the best of our knowledge, Hierarchical Interaction Message Passing mechanism[13, 14] and develop a unified molecular message passing network, HimNet. In our approach, atoms, motifs, molecules, and molecular fingerprints are treated as distinct semantic layers[18, 19]. Crucially, we introduce the Explainable Attention Interaction module[20, 21], which enables learnable modeling of interactions between functional groups. This design facilitates the capture of cross-motif cooperative mechanisms, including hydrogen bonding[22], π - π stacking[23], and hydrophobic effects[24]. Moreover, we incorporate a Consensus Fingerprint Enhancement module[19, 25] that identifies latent functional patterns via multi-fingerprint similarity analysis and fusion, thereby guiding global structural optimization. Finally, a multi-head attention fusion mechanism[20] aligns features across hierarchical levels, significantly improving both generalizability and interpretability. Extensive experiments on eleven datasets—including eight widely-used MoleculeNet benchmarks[3] and three challenging, high-value datasets for metabolic stability[26], malaria activity[27], and liver microsomal clearance[28]—demonstrate that HimNet consistently outperforms existing hierarchical GNN models[13, 14] and feature-fusion baselines[12], achieving best or near-best performance in most tasks.

2 Results

2.1 Hierarchical Interaction Message Passing Mechanism

To uncover the intricate dependencies that span multiple structural levels, HimNet incorporates a Hierarchical Interaction Message Passing Mechanism (HIMPM). Conventional message-passing schemes normally operate only on directly bonded atoms or on rigidly pre-defined fragments[6]. In contrast, real molecular systems often feature cooperative or competitive interactions that cross covalent boundaries—think of π - π stacking between adjacent aromatic rings[23] or the way a distant polar substituent modulates a hydrophobic cluster[24]. HIMPM shatters this single-view limitation by tying together atoms, motifs, and the molecule as a whole, enabling seamless information flow across scales and faithfully capturing these non-additive, multi-pathway behaviors[13, 14].

In practice, HIMPM runs two complementary pathways in parallel. One adopts a directed message-passing strategy (akin to D-MPNN)[7], using chemical bonds to precisely encode each atom’s local environment. The other leverages hierarchical interaction attention to erect “long-distance bridges” across atom-atom, atom-motif, motif-motif, and motif-global links, allowing the model to directly sense and weigh the influence between disparate structural fragments[20, 13]. At each iteration, a learnable gating mechanism fuses the outputs of these pathways, dynamically balancing strict chemical bond constraints with flexible, multi-scale semantic integration[13].

To better contextualize our approach’s strength in capturing long-range dependencies, we compare it with standard graph-based methods. Conventional GNNs (e.g., GCN, MPNN) propagate information via iterative neighbor aggregation, which is inefficient for connecting distant nodes in a molecular graph

and can lead to the over-smoothing problem. In contrast, our HIMP, through its hierarchical attention pathway, directly establishes "long-distance bridges" between different structural units (e.g., atoms and motifs), effectively bypassing topological bottlenecks in information flow. Furthermore, while attention-based methods like GAT or Graph Transformers also exist, they typically operate on a single, flat atomic level. HimNet is distinguished by its hierarchical attention, which operates across multiple semantic levels—atom, motif, and global—enabling a more precise modeling of the synergistic and antagonistic effects between functional groups that are key to determining complex molecular properties. This dual-pathway fusion of message passing and attention gives HimNet deep insight into intramolecular cooperation. Whether it is adjacent aromatic rings stacking cooperatively or hydrophobic and polar groups competing over several bonds, HIMP makes these effects explicit through its attention weights. As a result, HimNet not only achieves notable accuracy gains on tasks such as blood–brain barrier permeability, aqueous solubility, and toxicity prediction[3], but also provides clear, chemically grounded rationales for its decisions—substantially enhancing both performance and interpretability[21].

2.2 Model Overview: HimNet

Figure. 1b illustrates the overall architecture of HimNet, which integrates four tightly coupled stages to transform raw SMILES into a unified embedding for downstream prediction. In the first stage, Molecule Preprocessing, each SMILES string is parsed into a 2D chemical graph[29] and then fragmented by the BRICS algorithm[18] into atom-level nodes, motif-level nodes and a single global supernode. Explicit atom–motif, motif–motif and motif–global edges link these nodes, preserving fine-grained bond topology while exposing higher-order substructure relationships[13].

Running in parallel, Figure. 1b(ii) introduces the Fingerprint Embedding module, which encodes five classical molecular fingerprints—AtomPairs[30], MACCS keys[31], MorganBits and MorganCounts[19], and Pharmacophore[32]—into dense vectors. A pairwise cosine similarity matrix highlights high-confidence shared dimensions[33], which are extracted as "consensus" features and then fused with the original fingerprint embeddings via learnable attention weights and a contrastive alignment loss[34], thereby injecting global chemical priors that complement the graph encoding.

The third component is the Hierarchical Interaction Message Passing Mechanism (HIMP) shown in Figure. 1b(iii)[13, 14]. One path follows directed message passing (D-MPNN)[7] to rigorously propagate local atomic information along chemical bonds, while the parallel interaction-attention path builds long-distance bridges among atoms, motifs and the global node to capture non-additive cooperativity such as π – π stacking, hydrogen-bond networks and hydrophobic clustering[20]. Learnable gating weights adaptively fuse these two channels at each iteration, ensuring strict adherence to bond constraints together with flexible integration of multi-scale semantics.

Figure. 1b(iv) depicts the Multi-Head Attention Fusion module, which aligns and merges the multi-scale graph representations from HIMP with the consensus fingerprint embedding through data-driven attention weights[20]. The resulting compact, highly expressive vector drives downstream classifiers and regressors, achieving state-of-the-art accuracy on blood–brain barrier permeability, solubility and toxicity prediction[3]. Moreover, its attention maps provide clear, chemically grounded explanations that greatly enhance interpretability[21].

2.3 Accurate Molecular Property Prediction with HimNet

We first evaluated the performance of HimNet on eight classic benchmark datasets defined by MoleculeNet—including five binary classification tasks (BACE, BBBP, Tox21, SIDER, ClinTox) and three regression tasks (ESOL, FreeSolv, Lipophilicity)[3]—covering a broad range of physicochemical and bioactivity prediction scenarios. To further assess the applicability of HimNet in more advanced and realistic drug discovery contexts, we additionally included three less commonly studied but highly representative datasets: Malaria (anti-malarial EC_{50} regression)[27], LMC (liver microsomal clearance regression across species)[28], and MetStab (binary classification of metabolic stability)[26]. Although these three datasets

are not newly curated, they have been much less explored in the literature and present challenging tasks that better reflect the complexity encountered in real-world pharmaceutical research[14].

For all eleven datasets, we adopted scaffold-based splitting (80% training / 10% validation / 10% test) following the protocol in MoleculeNet[3]. Three independent splits were generated using different random seeds to ensure statistical robustness. Unlike random splitting, scaffold splitting groups molecules by their core scaffolds, minimizing the chance of similar chemical backbones appearing in both training and test sets, thus providing a more realistic and stringent assessment of model generalization[3]. For each split, we performed ten training runs with different random weight initializations, and all models were trained using the Adam optimizer for 100 epochs[35].

To ensure comprehensive and fair comparison, we selected a range of recent state-of-the-art molecular graph neural network baselines, covering hierarchical/multiscale modeling, geometry or image augmentation, and fingerprint integration strategies. Specifically, the baselines include:

- Hierarchical/multiscale GNNs (e.g., MolCLR[15], HiMol[13], HimGNN[36]), which represent information at atom, motif, and graph levels but often lack explicit and dynamic inter-level interaction;
- Geometry and image-enhanced methods (e.g., GEM[37], Uni-Mol[38], ImageMol[39]), which enrich graph representations with 3D or visual features, though the integration with chemical semantics is often shallow;
- Fingerprint-augmented GNNs (e.g., FH-GNN[14]), which combine traditional molecular fingerprints with learned graph features, but typically only at the final output stage.

HimNet addresses these limitations by introducing an end-to-end bidirectional hierarchical message passing mechanism that allows atom, motif, and graph-level representations to evolve jointly and interactively. Furthermore, HimNet employs an attention-based fusion of molecular fingerprints and learned features throughout the entire architecture, not merely at the output. This unified and flexible framework enables the model to dynamically balance local chemical detail and global molecular context, thus enhancing its capacity to model complex molecular properties.

The experimental results, summarized in Tables 1 and 2, show that HimNet achieves state-of-the-art or highly competitive performance on the majority of tasks. For example, HimNet obtains the highest ROC-AUC scores on BACE (0.890), BBBP (0.954), Tox21 (0.826), and ClinTox (0.950)[40]. In regression benchmarks, HimNet achieves the lowest RMSE on ESOL (0.710) and FreeSolv (1.441)[41], demonstrating its strong capability in modeling physicochemical properties.

To comprehensively evaluate the performance of HimNet on more challenging tasks, we further compared it against a series of baseline models on the MetStab, Malaria, and LMC datasets. Recognizing that these datasets are less commonly benchmarked in the literature, we deployed four representative graph neural networks (MPNN, GCN, GAT, GIN) and additionally included four advanced hierarchical models (FH-GNN, MolCLR, HiMol, HimGNN) for a more rigorous evaluation. As shown in Table 2, HimNet demonstrates exceptional performance on these tasks. On both the MetStab and Malaria datasets, HimNet’s metrics are significantly better than all competing hierarchical models. For the complex LMC multi-species clearance prediction task, HimNet also achieves a superior mean RMSE compared to FH-GNN, MolCLR, and HiMol. These results provide strong evidence of HimNet’s robust generalization capabilities and superiority in new and complex scenarios that more closely reflect real-world pharmaceutical research.

A dataset-wise analysis reveals that HimNet also performs robustly on the less commonly studied Malaria[27] (RMSE 1.029) and MetStab[26] (AUC 0.896) datasets, outperforming all baselines and highlighting its generalization strength in new and challenging tasks. On the LMC dataset[28], HimNet achieves the lowest average RMSE (111.0) across the three species, with particularly strong results for human (91.7) and rat (109.3). However, on the mouse subset (LMC-M), the model’s performance (RMSE 132.1) is slightly behind that of GAT (124.8), likely due to higher experimental noise and inter-species variability, which reduce the advantage of hierarchical modeling and make simpler models more robust to noise[14].

On the SIDER dataset, which involves multi-label classification with highly imbalanced and long-tailed label distribution, HimNet achieves an AUC of 0.651. Although competitive, this does not surpass some

baselines; the probable reason is that SIDER contains numerous rare adverse reaction labels, making it difficult for hierarchical models to learn effective signals, while approaches that rely more on global or late-stage feature fusion (such as FH-GNN) may better exploit statistical correlations in imbalanced data[14].

For Lipophilicity regression, HimNet reaches an RMSE of 0.698, which is close to but does not exceed the best baseline performance (HimGNN at 0.632). This may be attributed to the limited sample size and the nuanced nature of lipophilicity, which depends on both local functional groups and global molecular context. While HimNet’s hierarchical approach excels when sufficient training data is available to capture these dependencies, its effectiveness is somewhat reduced in low-data and context-sensitive scenarios.

In summary, HimNet delivers state-of-the-art or highly competitive performance on almost all molecular property prediction tasks, especially excelling on those where both hierarchical structure and chemical priors are crucial. Its relatively weaker results on datasets characterized by extreme label imbalance, small sample size, or high experimental noise (such as SIDER, Lipophilicity, and LMC-M) indicate the challenges that remain for hierarchical modeling in such settings. Future work could incorporate strategies such as sample rebalancing[42] or noise-aware training to further improve robustness and adaptability.

2.4 Explanation prediction

Understanding how molecular structure influences predictive attention is essential for interpreting the decision-making process of graph-based models[20]. To this end, we conducted a case study using the BBBP dataset, which contains annotated information on the blood–brain barrier permeability of a wide range of compounds[3]. This property is particularly relevant in the early stages of central nervous system (CNS) drug development, where identifying molecules capable of penetrating the barrier is critical for reducing potential toxicity and improving therapeutic efficacy[43].

The visualizations in Figure. 2 and Figure. 3 illustrate how different molecular substructures contribute to blood–brain barrier permeability as interpreted by the model’s attention mechanisms[21]. Regions with higher attention scores—depicted as brighter areas—are considered to have a greater influence on the model’s prediction of permeability[20]. In these visualizations, red-colored regions indicate molecular components that are predicted to facilitate BBB penetration, whereas blue-colored regions correspond to structural elements that are predicted to hinder it.

Analysis of Molecule 1: CC1COc2c(N3CCN(C)CC3)c(F)cc3c(=O)c(C(=O)O)cn1c23 Figure. 2a displays the distribution of attention weights at the atomic level within the hierarchical molecular graph. The hue and intensity of each atom indicate both the direction and magnitude of its contribution to BBB permeability as assessed by the model. Notably, red regions are concentrated on the piperazine ring and its adjacent carbon atoms on the left side of the molecule, as well as certain segments of the fused ring skeleton. From a chemical perspective, the piperazine ring, despite containing nitrogen atoms, exhibits a more aliphatic character compared to regions such as carboxyl groups[44]. This property increases the molecule’s overall lipophilicity, which may be attributed to enhanced permeability across the lipid-rich blood–brain barrier[43]. Thus, the model’s assignment of red coloration—indicating a positive contribution—is consistent with established chemical understanding.

Conversely, blue regions are predominantly found on the carboxyl ($-\text{COOH}$) group at the right side of the molecule, along with nitrogen and carbonyl oxygen atoms within the fused ring structure. The carboxyl group, being highly polar, is likely to dissociate into a negatively charged carboxylate ($-\text{COO}^-$) at physiological pH[45]. Its strong polarity, negative charge, and capacity for hydrogen bonding substantially increase the polar surface area and reduce lipophilicity[46], both factors known to hinder BBB penetration. Similarly, the presence of polar ring nitrogen and carbonyl oxygen atoms, which can form hydrogen bonds, further elevates the polar surface area, thereby impeding BBB permeability[47]. This suggests that the model’s attention allocation closely aligns with fundamental chemical principles[21].

Figure. 2b aggregates atomic contributions into larger chemical motifs or functional groups, illustrating their collective attention weights. In this visualization, the red region on the left corresponds to the

piperazine ring (M1), the blue region on the right represents the carboxyl group (M2), and the light red area in the center denotes the core fused ring scaffold (M0). The model thus identifies the piperazine motif as generally facilitating blood–brain barrier permeability, while the carboxyl motif acts as a barrier—findings that are consistent with the chemical rationale discussed above.

Figure. 2c provides an integrated, hierarchical perspective on attention allocation within the molecular graph. This visualization reveals that the model’s prediction of blood–brain barrier permeability is shaped not only by the intrinsic properties of individual motifs, but also by the patterns of interaction among them[21].

Crucially, these findings directly reflect the internal working mechanism of HimNet, which leverages hierarchical message passing across atomic, motif, and graph levels to encode chemically meaningful multi-scale dependencies[13, 14]. Within HimNet, atomic-level messages are first aggregated into motif-level representations via learnable, context-aware attention[20, 21], capturing the essential chemical environments of functional groups. Subsequently, motifs communicate with each other and with the graph-level node through cross-level attention pathways, allowing the model to explicitly encode the non-additive, context-dependent interactions between different structural components[20, 13].

This mechanism is fundamentally distinct from conventional node-level message passing: instead of simply summing local neighborhoods[48], HimNet’s hierarchical structure enables a motif to reference not only its own atoms but also other motifs and even remote atomic environments[13, 14]. For example, in this molecule, the positive effect of the fluorinated aromatic motif (M0) and the pyridine motif (M1) is not simply a result of their local features, but is contextually modulated and even amplified in the presence of the polar ester motif (M2)[15]. Conversely, the negative contribution of M2 is intensified when considered together with neighboring hydrophobic motifs, reflecting a chemically accurate picture where polarity and hydrophobicity are in competition[24].

This direct, multi-scale message exchange and attention aggregation empower HimNet to model higher-order cooperative and antagonistic effects (such as conjugation, hydrogen bonding networks, or lipophilic clustering)[22, 23, 24] that underlie complex molecular properties.

In practice, each motif node in HimNet acts both as an aggregator of atomic information and as an active agent in hierarchical interactions[13]. Motif-to-motif and motif-to-atom attention edges, which can extend beyond motif boundaries, provide a mechanism for encoding the chemical “fuzziness” of functional group effects[21]—allowing, for example, the ester motif (M2) to reference adjacent aromatic carbons (in M0), and vice versa[15]. These cross-hierarchy, cross-context interactions mean that property predictions arise not from linear combinations of isolated features, but from coordinated, emergent behavior across the molecular graph, providing a rigorous theoretical foundation for chemical interpretability[21].

A closer examination reveals that each motif node functions not only as an aggregator of its constituent atomic information, but also as an active participant in interactions with atoms outside its defined boundaries[20, 13]. For instance, the ester motif (M2) establishes links with aromatic carbon atoms in the fluorinated core (M0), thereby enhancing its disruptive polarity in an otherwise lipophilic environment. Conversely, M0 accesses information from carbonyl oxygen atoms in M2, potentially counteracting the negative effects of excessive polarity via local contextual adjustment[13]. These cross-motif and motif–atom interactions indicate that the model learns context-aware semantic representations for each structural unit, allowing it to assess the contribution of each motif within its broader chemical environment[20]. Ultimately, this demonstrates that molecular properties constitute necessary consequences of coordinated component interactions, rather than mere linear combinations of isolated motifs[21].

Analysis of Molecule 2: CCC(=O)C(CC(C)N(C)C)(c1ccccc1)(c1ccccc1) In the case of the second molecule, the atomic-level attention map displays a distinct distribution pattern. As depicted in Figure. 3a, regions of high attention—marked in red—are primarily located on the carbon atoms of the two benzene rings and parts of the central carbon framework. This pattern suggests that the model identifies these hydrophobic domains as positive contributors to blood–brain barrier permeability, aligning well with established chemical knowledge that lipophilic aromatic groups tend to facilitate BBB penetration[43].

Conversely, areas of blue coloration are centered around the nitrogen atom in the dimethylamino group ($-\text{N}(\text{CH}_3)_2$) and the oxygen atom in the carbonyl group ($\text{C}=\text{O}$). While tertiary amines may, in some contexts, promote BBB permeability, the model here assigns a negative contribution to this functional group, which may be attributed to unfavorable interactions with neighboring structural motifs[47]. The carbonyl oxygen is also assigned negative attention, consistent with its high polarity and capacity to serve as a hydrogen bond acceptor, both of which are known to impede BBB permeability[47].

The motif-level attention analysis in Figure. 3b highlights three principal functional groups, each characterized by a distinct attention weight. The first, the dimethylamino fatty chain (M1), receives a light blue score of 0.21. The ketone motif (M0) is similarly assigned 0.21, reflecting its polar character. Among the aromatic moieties, one phenyl ring (M2) is rendered in strong red with a score of 0.47, while the other ring (M3) appears in blue with a lower score of 0.10. This clear disparity between two structurally similar benzene rings demonstrates the model’s ability to differentiate aromatic systems based on their specific molecular environments and contextual interactions, rather than treating them as equivalent substructures[15].

Figure. 3c provides further insight into how the HimNet model’s hierarchical message passing mechanism enables the emergence of chemically meaningful, context-dependent interactions among structural motifs. Within this integrated view, each motif not only aggregates information from its constituent atoms but also exchanges messages with other motifs and, crucially, interacts with both adjacent and remote atomic environments across the molecule[13, 20]. The attention pathways learned by HimNet reflect a multi-scale theoretical framework, in which atom-level messages are first pooled into motif representations via attention, and subsequently, motifs propagate their semantic information to the global graph node through additional cross-level message passing[48].

A key manifestation of this mechanism is that the two phenyl motifs (M2 and M3), though structurally similar, are contextually distinguished: one receives strong positive attention, while the other is negatively weighted, depending on the influence of their spatial and electronic environment—such as proximity to polar carbonyl (M0) or amine (M1) motifs[15]. This non-additive and position-sensitive differentiation cannot be achieved by conventional graph networks relying solely on local aggregation, but arises naturally from HimNet’s explicit cross-hierarchy message flow[13].

Moreover, polar motifs (M0, M1) exert inhibitory effects not only through their intrinsic properties but also by influencing the contribution of nearby aromatic regions, mirroring competitive and cooperative interactions between hydrophobic and polar domains seen in real chemical systems[24, 22]. Motif-atom and motif-motif interactions further illustrate the model’s capability to capture “boundary effects” and extended chemical context, reflecting inductive, conjugative, or steric effects that propagate beyond simple substructure boundaries.

The patterns observed in both molecules thus exemplify the model’s ability to construct a hierarchical, context-aware representation of molecular structure—progressing from atom-level features to motifs, and ultimately to global property prediction[21]. HimNet’s layered message passing and explicit multi-scale aggregation capture not only local chemical attributes but also their broader contextual significance, ensuring that property prediction is grounded in the structured, emergent interactions among all relevant molecular components[13].

2.5 Ablation Study

To rigorously assess the contribution of each module in HimNet, we conducted ablation experiments by selectively removing key components and evaluating each variant on eight widely-used MoleculeNet benchmarks[3, 13]. These benchmarks span both classification and regression tasks and are well-established in the literature for systematic model analysis. The three additional datasets (Malaria, LMC, and Met-Stab), introduced in our main performance comparison to further demonstrate HimNet’s generalizability on challenging real-world endpoints, are not included in the ablation study. This choice ensures clarity and interpretability of the ablation results, as the selected benchmarks already provide a comprehensive and representative testbed for isolating the effect of each architectural component[14]. The ablation settings are as follows:

A summary of results is presented in Table 3 and visualized in Figure 4, where each row (table) and colored bar (figure) corresponds directly to a specific model variant.

Table 3 quantitatively reports the mean and standard deviation of performance metrics (e.g., ROC-AUC for classification, RMSE for regression) across three runs on each benchmark dataset[3]. Figure 4 provides a visual comparison, in which the performance of each ablated variant is shown alongside the full model. This enables a direct and intuitive understanding of the impact of each component.

Across all datasets, the full HimNet model consistently achieves the best or near-best results (bolded in the table, leftmost bars in the figure). Removing any individual module results in a noticeable performance drop across most tasks, as clearly seen in both the numerical values in Table 3 and the relative bar heights in Figure 4. In particular, removing the global-local attention (w/o I) or the consensus fingerprint enhancement (w/o C) leads to especially large decreases in both classification (e.g., Tox21, SIDER) and regression (e.g., FreeSolv, ESOL) datasets. This is evident in Table 3 by the lower scores in these columns, and in Figure 4 where the corresponding bars dip most significantly below the full model[13]. The necessity of multi-head attention fusion (w/o A) is most apparent in tasks such as ClinTox and Lipophilicity, where its removal causes a marked reduction in performance (also highlighted in both the table and figure). Even the core hierarchical message passing modules (w/o M and w/o H) remain foundational, as their ablation consistently results in less accurate predictions across all datasets. A comprehensive assessment of model complexity, training duration, and inference speed is essential for real-world applications, particularly in large-scale virtual screening. We conducted a comparative evaluation of HimNet against two other hierarchical architectures, FH-GNN and HIMoI, on the ClinTox and Freesolv datasets. As the experimental results in Table 4 demonstrate, a compelling trade-off is revealed: HimNet attains significantly stronger predictive performance with a more lightweight parameter footprint, yet it requires more computational time for both training and inference than its counterparts. In summary, the ablation results—numerically detailed in Table 3 and visually corroborated by Figure 4—demonstrate that each module in HimNet plays a vital and complementary role. Furthermore, the comparative analysis presented in Table 4 confirms that our model achieves a favorable balance between predictive accuracy and parameter efficiency against other hierarchical architectures. The systematic drop in performance upon any module’s removal validates that these architectural components are indispensable for robust and accurate molecular property prediction[14].

As shown in Figure 5, a comprehensive ablation study was conducted to visually illustrate the impact of each fingerprint and their combinations. Figure 5a depicts single-fingerprint ablations and visually underscores the task-dependent importance of each feature set. For example, the relatively low ROC-AUC bar for the SIDER dataset within the `morgan_counts` subplot, when compared to its counterparts in other subplots, visually confirms the critical role of this fingerprint in that specific task. Furthermore, Figure 5 also provides strong visual evidence for the synergistic effects among fingerprints. This effect is most apparent when comparing the paired-ablation results presented in Figure 5b to the single-ablation results in Figure 5a. A compelling visualization of this synergy can be observed on the CLINTOX dataset. The ROC-AUC bar in the `morgan.bits+pharmacophore` subplot is markedly lower than the corresponding bars in both the individual `morgan.bits` and `pharmacophore` subplots. This visual comparison reinforces that the combined removal of two fingerprints often leads to a disproportionately larger performance drop than removing either one alone. In aggregate, the widespread performance degradation shown across the entire grid of plots robustly supports our main conclusion. The strength of our model lies in the effective integration of a diverse and multifaceted set of chemical fingerprints, whose contributions are both task-dependent and highly synergistic.

As visually summarized in Figure 6, we conducted a comprehensive ablation study to elucidate the relative importance of different hierarchical interactions. The figure illustrates the performance of five model variants, each lacking a specific interaction mechanism. This visualization clearly demonstrates that the contribution of each interaction pattern is highly task-dependent, reinforcing the rationale for their integration in a multifaceted model.

The critical role of certain interactions for specific tasks is visually apparent. For instance, the subplot for CLINTOX shows that the `atom-atom` bar is markedly lower than all others, signifying that the removal

of atomic-level message passing leads to the most severe performance degradation for this task. This underscores the fundamental importance of local atomic information in predicting clinical toxicity. Conversely, for regression tasks, a different pattern emerges. The subplots for **ESOL** and **Lipophilicity** reveal that the **atom-motif** bar is the highest among all RMSE values, indicating that ablating the interaction between atoms and motifs is most detrimental. This finding suggests that capturing relationships between local atomic environments and larger, chemically meaningful motifs is essential for accurately predicting these physicochemical properties. Notably, on the **FreeSolv** task, removing certain complex interactions, especially motif-graph, led to improved performance. This reveals a deeper insight: for some tasks with relatively simple features, an overly complex interaction model may capture noise or lead to overfitting. In summary, Figure 6 provides compelling visual evidence that no single interaction is universally dominant or dispensable.

2.6 Hyperparameter Sensitivity Analysis

To assess the robustness of our proposed model, HimNet, and validate our hyperparameter selections, we conducted a comprehensive sensitivity analysis. We systematically investigated the model’s performance in response to variations in four critical hyperparameters: encoder depth, learning rate, hidden size, and the number of attention heads.

Our methodology involved varying one hyperparameter at a time while holding others constant at their optimized default values. For each setting, the model was trained and evaluated across all eight benchmark datasets, with performance averaged over three independent runs. The results are visualized in Figure 7.

Overall, HimNet exhibits remarkable robustness across a practical range of hyperparameter configurations:

- **Encoder Depth.** As shown in Figure 7a, the model’s performance is highly stable for encoder depths ranging from 3 to 9. Notably, several regression tasks, such as **ESOL** and **Lipophilicity**, achieve a slight performance peak at a depth of 7, validating our default choice.
- **Learning Rate.** The results in Figure 7b indicate that this parameter displayed the most pronounced sensitivity. While a higher rate of 0.001 led to a discernible performance degradation on datasets such as **SIDER** and **Lipophilicity**, the range from 5×10^{-5} to 5×10^{-4} consistently yielded strong results. This suggests that the latter range represents an optimal trade-off for balancing performance across diverse tasks. Our final choice of 1×10^{-4} falls within this robust region, ensuring stable and reliable performance.
- **Hidden Size.** According to Figure 7c, HimNet is largely insensitive to variations in hidden size within the 256 to 1024 range, maintaining strong and consistent performance. Conversely, a smaller dimension of 128 resulted in a noticeable decline on certain tasks, thus justifying the selection of a larger model capacity to capture complex molecular interactions.
- **Number of Heads.** Figure 7d demonstrates the model demonstrates exceptional structural robustness, with performance being nearly invariant to the number of attention heads for values between 4 and 16. This finding suggests that the hierarchical interaction mechanism is not overly dependent on a precise number of heads, enhancing its practical applicability.

In summary, this analysis establishes the robustness of HimNet, demonstrating that it does not require exhaustive fine-tuning to achieve high performance. This inherent stability validates the rationale behind our hyperparameter selection and strongly suggests the model’s capacity for generalization to new molecular prediction tasks. Thus, we present the best hyperparameter our model use in Table 5.

2.7 Model Complexity and Efficiency Analysis

To evaluate the practical potential of HimNet, we analyzed its computational efficiency in comparison to HiMol and FH-GNN under a unified environment. The results in Table 4 indicate that HimNet is

the most optimal in terms of parameter count, demonstrating higher parameter efficiency. However, its training and inference times are longer than the other two models, a fact we attribute to its more complex cross-level attention and gating fusion mechanisms. We argue that in the context of drug discovery, this increased computational cost is a justifiable trade-off for superior predictive performance. A more accurate model can more reliably prioritize candidate compounds, potentially saving substantial costs associated with expensive experimental validation. Therefore, while HimNet is not superior in training speed, its significant improvement in predictive accuracy makes it a more valuable and compelling choice for practical applications.

3 Discussion

3.1 Research Findings and Contributions

This study presents HimNet, a hierarchical molecular graph neural network grounded in the principle of hierarchical interaction learning[13]. By explicitly modeling interactions across atomic, motif, and molecular graph levels, HimNet captures non-additive synergistic effects between functional groups, effects often neglected by conventional GNN architectures[14]. This multiscale interaction mechanism improves the model’s ability to represent complex structure-property relationships, thus improving both predictive precision and interpretability, as confirmed by benchmark evaluations (Section 2, Tables 1, 2).

A central innovation of HimNet lies in its dual-pathway message passing architecture, which, in conjunction with multi-level attention, enables the encoding of cross-scale, nonlinear dependencies and functional group cooperativity—critical features for accurate molecular property prediction[20]. Complementing this architecture, a consensus fingerprint fusion module is introduced to adaptively extract and integrate shared chemical information from diverse fingerprint types[19]. This enrichment of molecular representation facilitates alignment between global and local semantic features. Furthermore, HimNet leverages multi-head attention to align fingerprint-based and graph-based representations at different levels, promoting both generalization and interpretability in its predictive outputs.

Empirical results demonstrate that HimNet consistently outperforms established baselines in accuracy and generalizability[3]. Beyond performance metrics, ablation experiments and visualization analyses highlight that explicitly capturing hierarchical cooperative effects leads not only to improved predictive power but also to chemically meaningful interpretability of model behavior[21].

3.2 Applicability and Limitations

HimNet offers a general and effective framework for modeling multi-scale molecular interactions, with particular advantages in scenarios where molecular properties are influenced by the cooperative behavior of substructures—such as in small molecules, drug-like compounds, and ligands of moderate topological complexity.

Nonetheless, the framework exhibits limitations in handling molecules with high conformational flexibility, complex stereochemistry, or macrocyclic structures. In such cases, the current BRICS-based decomposition[18] and two-dimensional graph representation[29] may fall short, as they do not fully capture essential three-dimensional conformational effects or dynamic interactions. Properties like conformational entropy, chiral recognition, and shape-dependent permeability often depend on these spatial characteristics, underscoring the need for incorporating 3D structural information or alternative decomposition strategies in future developments.

Moreover, while attention mechanisms in HimNet provide a degree of interpretability by highlighting influential atoms or motifs, it is important to acknowledge their limitations: attention weights, although informative, do not inherently reflect causal relationships and should be interpreted as heuristic guides rather than definitive indicators of feature importance.

3.3 Addressing Limitations

To mitigate the outlined limitations, several practical improvements can be pursued. First, enhancing computational efficiency through model compression or streamlined attention mechanisms would enable scaling HimNet to larger datasets without compromising performance. Second, data augmentation and robustness techniques, such as SMILES enumeration, graph perturbations, or self-supervised pretraining[15], can improve generalization, particularly in settings with limited or imbalanced data. Third, to bolster interpretability, attention-based methods can be complemented with more rigorous explainable AI approaches—such as Shapley value analysis or counterfactual explanations—which provide deeper insights into model decision-making[21].

3.4 Future Research Directions

Building on the current work, several avenues merit further investigation. Integrating 3D geometric and stereochemical information into the hierarchical representation—as demonstrated in Uni-Mol[38]—would extend the model’s applicability to flexible and chiral molecules. Additionally, scaling HimNet to support large-scale datasets and high-throughput screening applications, or adapting it for transfer learning in related domains (e.g., protein–ligand interaction prediction or materials informatics), offers substantial promise. Finally, incorporating multimodal molecular data, including protein structures, experimental spectra, or molecular dynamics trajectories, could establish a comprehensive framework for property prediction that captures the full complexity of real-world chemical systems.

4 Conclusion

In this work, we have addressed the critical challenge of accurately predicting small-molecule properties—particularly ADMET profiles—in modern drug discovery by introducing a *Hierarchical Interaction Message Passing Mechanism* at the core of our model, HimNet. Unlike existing deep learning approaches (e.g., GNNs[6] and Transformers[20]) that treat molecular graphs in a largely flat or fragment-centric manner, HimNet explicitly models atoms, motifs, and whole-molecule fingerprints as interlinked semantic layers. Through hierarchical attention-guided message passing, our framework enables dynamic, interaction-aware representation learning that balances global topology and local chemical environments. Extensive evaluation on benchmark datasets—including classic tasks such as BBBP, Tox21, and ESOL[3], as well as three challenging, high-value datasets for metabolic stability (MetStab)[26], malaria activity (Malaria)[27], and liver microsomal clearance (LMC)[28]—demonstrates that HimNet achieves state-of-the-art or near-state-of-the-art performance across diverse molecular property prediction tasks. Moreover, the built-in Explainable Attention Interaction[21] and Consensus Fingerprint Enhancement modules grant HimNet a strong degree of hierarchical interpretability, aligning model attributions with established chemical intuition on representative compounds. These results confirm that explicitly capturing cross-layer functional-group synergies is key to improving both predictive accuracy and mechanistic insight.

Despite these advances, several avenues remain for further improvement. First, extending HimNet to incorporate three-dimensional conformational data and solvent effects (e.g., as in Uni-Mol[38]) may enhance its ability to predict stereospecific and environment-sensitive properties. Second, reducing the computational overhead of multi-layer attention mechanisms will be essential for large-scale virtual screening campaigns. Finally, integrating our hierarchical paradigm into multi-task or multi-omics frameworks could unlock deeper insights into polypharmacology and mechanism-of-action inference.

In summary, HimNet offers an accurate, efficient and interpretable solution for molecular activity and ADMET prediction, thereby contributing significantly to informed decision-making in the early stages of drug discovery. Future work will focus on scaling and generalizing this hierarchical interaction approach to meet the evolving demands of small-molecule design and optimization.

5 Methods

This section details our proposed HimNet model from three aspects: first, the construction of hierarchical molecular graphs; second, the core mechanisms including the Hierarchical Interaction Message Passing Network (HIMPM), molecular fingerprint encoding module, and attention fusion; finally, the experimental setup and parameter configurations.

5.1 Hierarchical Molecular Graph Construction

To comprehensively capture molecular features and interactions at different scales, we constructed a three-level hierarchical graph representation including atom-level, motif-level, and global-level[13, 14]. This multi-scale representation simultaneously encodes local chemical details and complex topological relationships among molecular substructures.

Formally, the hierarchical molecular graph is defined as $G = (V, E)$, where the node set $V = \{V_a, V_m, V_g\}$ consists of atom nodes V_a , motif nodes V_m , and a global node V_g [13]. The edge set $E = \{E_a, E_m, E_{am}, E_{mg}\}$ comprises atom-atom bonds E_a , motif-motif relationships E_m , atom-motif associations E_{am} , and motif-global connections E_{mg} .

The construction process is illustrated in Figure 8. The process begins by parsing the molecular topology from SMILES strings to obtain atom nodes and chemical bonds, as shown in Figure 8c. Next, we apply the BRICS (Breaking of Retrosynthetically Interesting Chemical Substructures) decomposition algorithm to identify functional groups, generating the motif node set displayed in Figure 8b. Based on the decomposition results, we establish atom-motif associations by creating an edge between each motif and all of its constituent atoms. Simultaneously, motif-motif connections are formed as depicted in Figure 8d: an edge is created between two motif nodes if their respective constituent atoms were linked by a chemical bond prior to decomposition. Finally, we introduce a virtual global node, which acts as an information aggregation hub, and establish connections between it and all motif nodes. This completes the hierarchical molecular graph, with its final structure presented in Figure 8e. This hierarchical graph structure concurrently captures both local chemical environments and global structural features.

For feature initialization, atom node features include atom type, hybridization state, formal charge, and chemical environment information[48, 3]; motif node features are obtained by aggregating their constituent atoms’ features; and the global node feature is formed by aggregating all motif features[13].

This structured hierarchical representation lays the foundation for subsequent message passing and feature fusion.

5.2 Hierarchical Interaction Message Passing Network (HIMPM)

The Hierarchical Interaction Message Passing Mechanism (HIMPM) is the core innovation of our model, designed specifically to address the limitations of existing hierarchical GNNs by grounding its architecture in chemical first principles. Unlike models that rely on static, topology-based connections (e.g., HiMol, FH-GNN), which cannot capture non-local functional correlations, or those that employ late-stage fusion (e.g., HimGNN), which overlook the continuous interplay between different scales during learning, HIMPM introduces two key theoretical advancements. First, it employs a dynamic cross-level attention mechanism, breaking free from physical connectivity to learn a "functional interaction graph." This allows any atom or motif to directly interact with any other, enabling the model to capture long-range synergistic effects, such as conjugation, that are fundamental to determining molecular properties. Second, HIMPM features a deep, layer-wise gated fusion of two channels: a D-MPNN pathway for local bond-centric information and a hierarchical interaction attention pathway for global, multi-scale context. By integrating these channels at each update step via an adaptive gate, HIMPM ensures that the representation of each node continuously evolves under the dual influence of its local chemical environment and the global functional landscape. This process more faithfully simulates the emergent nature of molecular properties, where local and global effects are inextricably and dynamically coupled throughout the entire molecule.

As shown in Figure.1a, HIMPM consists of three organically integrated functional modules: an innovative message passing module, a multi-scale interaction module, and a global-local attention module[13]. The innovative message passing module transcends the limitations of traditional GNN message passing by adopting a dual-channel architecture that deeply integrates directed message passing with hierarchical interaction attention mechanisms[7, 5]. In the D-MPNN base channel, messages are propagated along directed chemical bonds from source node v to target node w [7]:

$$\mathbf{m}_{vw}^{(t)} = \text{Message}(\mathbf{h}_v^{(t-1)}, \mathbf{h}_w^{(t-1)}, \mathbf{e}_{vw}) \quad (1)$$

\mathbf{m}_{vw}^t denotes the message vector transmitted from the source node v to the target node w at step t . This message is generated by our message function, $\text{Message}(\cdot)$ which takes as input the hidden states of the involved nodes from the previous layer \mathbf{h}_v^{t-1} for the source and \mathbf{h}_w^{t-1} for the target—along with the feature vector of the intervening edge \mathbf{e}_{vw} .

The aggregated message for node v is obtained by summing the information from all adjacent edges[48]:

$$\mathbf{m}_v^{(t)} = \sum_{w \in \mathcal{N}(v)} \mathbf{m}_{vw}^{(t)} \quad (2)$$

where \mathbf{m}_v^t is the aggregated message vector for node v . This process is executed iteratively for T time steps, allowing information to propagate and enrich across the graph. Upon completion of the final time step, T , the resulting aggregated message vector encapsulates the complete information that node v has gathered from its local bond topology. We formally denote this final output of the D-MPNN channel as $\mathbf{m}_v^{\text{D-MPNN}}$.

Unlike traditional message passing paradigms, we introduce a hierarchical interaction attention channel that not only captures long-range dependencies between nodes, but more importantly, enables structure-aware information exchange across hierarchies[5, 13]. These two channels are not simply running in parallel, but are deeply integrated through an innovative learnable gating fusion mechanism[49]:

$$\mathbf{m}_v^{(l)} = \alpha_v \cdot \mathbf{m}_v^{\text{D-MPNN}} + (1 - \alpha_v) \cdot \mathbf{m}_v^{\text{HIAAtt}} \quad (3)$$

where $\alpha_v = \sigma(W_{\text{gate}}[\mathbf{m}_v^{\text{D-MPNN}} || \mathbf{m}_v^{\text{HIAAtt}}])$ is an adaptive gating parameter, and $\mathbf{m}_v^{\text{D-MPNN}}$ represents the message from local topological bonds, while $\mathbf{m}_v^{\text{HIAAtt}}$ represents the message from the Hierarchical Interaction Attention mechanism capturing long-range dependencies. The resulting $\mathbf{m}_v^{(l)}$ is the final, adaptively fused message vector used to update node v at layer l .

It is crucial to emphasize that this dual-channel, parallel architecture is a core design principle of HIMPM, intended to synergistically integrate both local and non-local molecular information. In this design, the atom-, motif-, and global-level graph representations interact at the feature level, rather than being explicitly fused at the structural level. To effectively suppress redundancy and noise during cross-hierarchy information passing, we have incorporated a multi-level control mechanism. First, within the local information pathway, the DMPNN, through its directed, edge-based message passing paradigm, inherently reduces trivial information feedback, forming the first line of defense against redundancy. Second, when handling cross-level and long-range dependencies, the Hierarchical Interaction Attention (HIAAtt) acts as a dynamic information filter via its learned weights. Ultimately, the gating fusion mechanism α_v in Equation (3) serves as an adaptive arbitrator, dynamically balancing the contributions of the two pathways to ensure a synergistic integration of local chemical topology constraints and cross-scale semantic information. The efficacy of this design is empirically validated in our ablation study (see Section 3.5), which confirms its necessity and robustness for achieving superior model performance.

It is crucial to clarify that our approach employs an explicit fusion strategy at the graph construction stage. Specifically, the atom-, motif-, and global-level nodes are integrated into a single, structured hierarchical molecular graph before being input to the network (see Section 2.1 and Figure 1). Therefore, the subsequent message passing does not operate on parallel graphs, but rather unfolds within this unified, explicitly fused structure.

The multi-scale interaction module, as the core unit of HIMPM, systematically models three key interaction patterns through carefully designed multi-head interaction attention mechanisms[20, 13]: atom-

atom interactions capture local chemical environments and long-range electronic effects[23, 24]; motif-motif interactions encode cooperative actions between higher-order functional groups[18, 14]; and atom-motif interactions integrate microscopic chemical properties with macroscopic substructures[13]. For example, the interaction between atom node a and other atom nodes i is:

$$h_a^{\text{new}} = \sum_{i \in \mathcal{N}(a)} \alpha_{ai} \cdot \text{MultiScaleInteraction}(h_a, h_i) \quad (4)$$

Here, `MultiScaleInteraction` is our specially designed interaction function that adaptively adjusts interaction patterns according to the representation spaces of nodes at different hierarchies[20, 13]. Similarly, the bidirectional interaction between motif node m and atom node a is given by:

$$h_m^{\text{new}} = \sum_{a \in \mathcal{N}(m)} \alpha_{ma} \cdot \text{MultiScaleInteraction}(h_m, h_a) \quad (5)$$

$$h_a^{\text{new}} = \sum_{m \in \mathcal{N}(a)} \alpha_{am} \cdot \text{MultiScaleInteraction}(h_a, h_m) \quad (6)$$

These formulations extend conventional attention to heterogeneous node types and dynamically adjust interaction modes based on hierarchical position[5, 15], enabling HIMPMP to capture complex cross-scale synergistic effects.

The global-local attention module, as the highest-level component of HIMPMP, achieves deep integration of global molecular representations with local structural features. This module employs a bidirectional attention mechanism that allows the global node to selectively aggregate critical information from both atom and motif levels[5, 20]:

$$H_g^{(a)} = W_g \cdot \sum_{a \in \mathcal{N}(g)} \alpha_{ga} \cdot h_a \quad (7)$$

$$H_g^{(m)} = W_g \cdot \sum_{m \in \mathcal{N}(g)} \alpha_{gm} \cdot h_m \quad (8)$$

The final global representation integrates multi-level information through a carefully designed fusion function:

$$H_g^{\text{final}} = \text{GL-Fusion}(\{H_g^{(a)}, H_g^{(m)}\}) \quad (9)$$

Through this hierarchical fusion architecture, HIMPMP breaks the barriers of information exchange between layers in traditional GNNs, achieving seamless information flow from atoms to motifs to global molecular representations, greatly enhancing the model’s ability to model complex molecular structures and properties[13, 14].

5.3 Molecular Fingerprint Encoding Module

As shown in Figure.1b(ii), we designed a multi-modal molecular fingerprint encoding module that integrates five complementary molecular fingerprints: Atom Pairs (AtomPairs)[30], MACCS keys[31], Morgan Bits (MorganBits)[19], Morgan Counts (MorganCounts)[19], and Pharmacophore fingerprints[32]. Each fingerprint is processed through an independent encoder to obtain a fixed-dimensional representation:

$$F_i = \text{Encoder}_i(FP_i) \quad (10)$$

To extract common information between different fingerprint types, we calculate the cosine similarity matrix $S_{i,j}$ between fingerprint encodings[33] and extract shared features based on similarity:

$$\text{CommonFeature}_{i,j} = \left(\frac{F_i + F_j}{2} \right) \odot \mathbb{I}(\text{ElementSim}_{i,j} > \tau) \quad (11)$$

where $\text{ElementSim}_{i,j} = \hat{F}_i \odot \hat{F}_j$ represents element-wise similarity of normalized vectors.

The shared features are aggregated via a weighted sum, where the weights are derived from the pairwise cosine similarities, and then combined with the original fingerprint features:

$$F_{\text{final}} = \phi(W_f[F_{\text{weighted}}; F_{\text{enhanced}}]) \quad (12)$$

where $F_{\text{weighted}} = \sum_{i=1}^5 \alpha_i \cdot F_i$ is a dynamically weighted combination of fingerprints, and F_{enhanced} represents enhanced shared features[19].

5.4 Attention Fusion

As illustrated in Figure.1b(iv), the attention fusion module integrates features from HIMPM-generated molecular graph representations and fingerprint encoding through a multi-head self-attention mechanism[20]. Each attention head contains three linear transformations: Query (Q), Key (K), and Value (V), enabling feature interactions across different representation spaces:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (13)$$

The outputs of the multi-head attention are concatenated and linearly transformed to obtain the final fused representation:

$$X_{\text{fused}} = W_O[\text{head}_1; \text{head}_2; \dots; \text{head}_n] \quad (14)$$

This module achieves complementary feature fusion, effectively integrating the structural interaction information captured by HIMPM and the chemical property information encoded by molecular fingerprints, forming a comprehensive molecular representation for downstream task prediction[20].

5.5 Experimental Settings

We evaluated our HimNet model on a total of eleven datasets: eight widely used benchmark datasets, including five classification tasks and three regression tasks (BBBP, BACE, Tox21, SIDER, ClinTox, ESOL, FreeSolv, and Lipophilicity)[3]—as well as three challenging high-value datasets introduced in this work: Malaria (antimalarial activity regression)[27], LMC (liver microsomal clearance regression)[28], and MetStab (metabolic stability classification)[26]. Table 6 summarizes the basic information of each dataset.

To ensure rigorous evaluation of model performance, we adopted scaffold-based splitting with an 8:1:1 ratio for training, validation, and test sets across all datasets, following the protocol in MoleculeNet[3]. This splitting method ensures that test molecules have scaffold structures different from those in the training set, effectively assessing the model’s generalization capability. For classification tasks, we used the ROC-AUC metric[40]; for regression tasks, we evaluated performance using the root mean square error (RMSE)[41].

The HimNet model was implemented using PyTorch 1.12.0[50] and PyTorch Geometric 2.2.0[51], and trained on NVIDIA GPU environments (CUDA 11.3). Molecular operations and fingerprint generation were implemented using the RDKit 2022.03.5 library[52], with BRICS decomposition for motif identification[18] and five types of molecular fingerprints (AtomPairs[30], MACCS keys[31], Morgan Bits & Counts[19], and Pharmacophore[32]) for feature extraction. All experiments on all datasets followed the same preprocessing pipeline and evaluation protocol to ensure fair comparison with baseline methods. Our choice of ROC-AUC and RMSE aligns with the standard evaluation protocols of prominent benchmarks such as MoleculeNet, ensuring a fair and direct comparison with previously published state-of-the-art methods.

6 Author contributions

Huiyang Hong conceived and designed the study, performed experiments, and drafted the manuscript. Xinkai Wu assisted with experiments and manuscript writing. Hongyu Sun assisted with experiments

and revised the model code. Chaoyang Xie contributed to the initial model design and discussions on innovative ideas. Yuquan Li and Qi Wang supervised the project and contributed to all aspects of the research and manuscript.

7 Competing interests

Competing Interests The authors declare no competing interests. All authors have reviewed the statement and confirm that no financial or non-financial competing interests exist regarding the research and its results presented in this manuscript.

8 Acknowledge

This research was supported by National Key RD Program of China(2024YFD2001100, 7582024YFE0214300), National Natural Science Foundation of China (No. 62162008, 32560689), Natural Science Special Fund of Guizhou University (No.202409), Guizhou Provincial Science and Technology Projects ([2024]002, CXTD[2023]027), Guizhou Province Youth Science and Technology Talent Project ([2024]317), Guiyang Guian Science and Technology Talent Training Project ([2024] 2-15). We thank Public Big Data Supercomputing Center and National Key Laboratory of Green Pesticide of Guizhou University for providing high-performance computing resources.

9 Data availability

The datasets used in this study are publicly available and can be accessed at <https://github.com/Hugh415/HimNet>. All data generated or analysed during this study are included in this published article and its Supplementary Data, and please see Supplementary Data 1.

10 Code availability

The custom code for the Hierarchical Interaction Message Net (HimNet) is deposited in the Zenodo repository at <https://doi.org/10.5281/zenodo.18030100>. The code is also available on GitHub at <https://github.com/Hugh415/HimNet> under an MIT license.

References

- [1] J. A. DiMasi, H. G. Grabowski, and R. W. Hansen. Innovation in the pharmaceutical industry: New estimates of r&d costs. *Journal of Health Economics*, 47:20–37, 2016.
- [2] Leonardo LG Ferreira and Adriano D Andricopulo. Admet modeling approaches in drug discovery. *Drug discovery today*, 24(5):1157–1165, 2019.
- [3] Z. Wu, B. Ramsundar, E. N. Feinberg, J. Gomes, C. Geniesse, A. S. Pappu, K. Leswing, and V. Pande. Moleculenet: a benchmark for molecular machine learning. *Chemical Science*, 9(2):513–530, 2018.
- [4] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019.
- [5] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.

-
- [6] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. Neural message passing for quantum chemistry. *Proceedings of the 34th International Conference on Machine Learning*, 70:1263–1272, 2017.
- [7] K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea, A. Palmer, V. Settels, T. Jaakkola, K. Jensen, and R. Barzilay. Analyzing learned molecular representations for property prediction. *Journal of Chemical Information and Modeling*, 59(8):3370–3388, 2019.
- [8] Liang Song, Xin Zhan, and Rui Peng. Contextual message passing neural network for molecular property prediction. *Bioinformatics*, 37(24):4753–4759, 2021.
- [9] Cailum MK Stienstra, Liam Hebert, Patrick Thomas, Alexander Haack, Jason Guo, and W Scott Hopkins. Graphormer-ir: Graph transformers predict experimental ir spectra using highly specialized attention. *Journal of chemical information and modeling*, 64(12):4613–4629, 2024.
- [10] Huihui Liu, Bingjie Zhu, Shuyang Nie, Haoran Li, Yugang Lin, Tianyi Ma, Xin Shao, Qian Chen, Minjie Shen, Yanrong Zheng, et al. Advancing admet prediction through multiscale fragment-aware pretraining with msformer-admet. *Briefings in Bioinformatics*, 26(5):bbaf506, 2025.
- [11] Zhichao Tan, Youcai Zhao, Tao Zhou, and Kunsen Lin. Hi-mgt: a hybrid molecule graph transformer for toxicity identification. *Journal of Hazardous Materials*, 457:131808, 2023.
- [12] Minghui Li, Zikang Guo, Yang Wu, Peijin Guo, Yao Shi, Shengshan Hu, Wei Wan, and Shengqing Hu. Vidta: Enhanced drug-target affinity prediction via virtual graph nodes and attention-based feature fusion. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 42–47. IEEE, 2024.
- [13] Xuan Zang, Xianbing Zhao, and Buzhou Tang. Hierarchical molecular graph self-supervised learning for property prediction. *Communications Chemistry*, 6(1):34, 2023.
- [14] Shuo Liu, Mengyun Chen, Xiaojun Yao, and Huanxiang Liu. Fingerprint-enhanced hierarchical molecular graph neural networks for property prediction. *Journal of Pharmaceutical Analysis*, page 101242, 2025.
- [15] 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.
- [16] Bruna DP Costa, João CS Simões, Susana MM Lopes, Mafalda Laranjo, Ana Clara B Rodrigues, Ana Cristina Gonçalves, J Sérgio Seixas de Melo, M Filomena Botelho, Marta Pineiro, and Teresa MVD Pinho e Melo. Hydrazone-functionalized trans-a2b-corroles: effective synergy in photodynamic therapy of lung cancer. *Journal of Medicinal Chemistry*, 67(24):21934–21951, 2024.
- [17] Kamal M Dawood and Sobhi M Gomha. Synthesis and anti-cancer activity of 1, 3, 4-thiadiazole and 1, 3-thiazole derivatives having 1, 3, 4-oxadiazole moiety. *Journal of Heterocyclic Chemistry*, 52(5):1400–1405, 2015.
- [18] J. Degen, C. Wegscheid-Gerlach, A. Zaliani, and M. Rarey. On the art of compiling and using ‘drug-like’ chemical fragment spaces. *ChemMedChem*, 3(10):1503–1507, 2008.
- [19] D. Rogers and M. Hahn. Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*, 50(5):742–754, 2010.
- [20] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pages 5998–6008, 2017.

- [21] Zhaoping Xiong, Dingyan Wang, Xiaohong Liu, Feisheng Zhong, Xiaozhe Wan, Xutong Li, Zhaojun Li, Xiaomin Luo, Kaixian Chen, Hualiang Jiang, 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.
- [22] G. A. Jeffrey. An introduction to hydrogen bonding. 1997. doi:10.1021/ed076p759.1.
- [23] C. A. Hunter and J. K. M. Sanders. The nature of π - π interactions. *Journal of the American Chemical Society*, 112(14):5525–5534, 1990.
- [24] D. Chandler. Interfaces and the driving force of hydrophobic assembly. *Nature*, 437(7059):640–647, 2005.
- [25] J.-L. Faulon, C. J. Churchwell, and D. P. Visco. The signature molecular descriptor. 1. extending to generic chemical reactions. *Journal of Chemical Information and Computer Sciences*, 43(3):707–720, 2003.
- [26] S. Podlowska and R. Kafel. Metstabon—online platform for metabolic stability predictions. *International Journal of Molecular Sciences*, 19(4):1040, 2018.
- [27] F. J. Gamo, L. M. Sanz, J. Vidal, and et al. Thousands of chemical starting points for antimalarial lead identification. *Nature*, 465(7296):305–310, 2010.
- [28] J. Brian Houston and D. J. Carlile. Prediction of hepatic clearance from microsomes, hepatocytes, and liver slices. *Drug Metabolism Reviews*, 29(4):891–922, 1997.
- [29] D. Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 28(1):31–36, 1988.
- [30] R. E. Carhart, D. H. Smith, and R. Venkataraghavan. Atom pairs as molecular features in structure-activity studies: Definition and application. *Journal of Chemical Information and Computer Sciences*, 25(2):64–73, 1985.
- [31] J. L. Durant, B. A. Leland, D. R. Henry, and J. G. Nourse. Reoptimization of mdl keys for use in drug discovery. *Journal of Chemical Information and Computer Sciences*, 42(6):1273–1280, 2002.
- [32] Johnathan W Campbell and Konstantinos D Vogiatzis. Augmenting maccs keys with persistent homology fingerprints for protein–ligand binding classification. *Journal of Chemical Information and Modeling*, 65(15):8113–8126, 2025.
- [33] David J Wood, Jacob de Vlieg, Markus Wagener, and Tina Ritschel. Pharmacophore fingerprint-based approach to binding site subpocket similarity and its application to bioisostere replacement. *Journal of chemical information and modeling*, 52(8):2031–2043, 2012.
- [34] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A simple framework for contrastive learning of visual representations. *International Conference on Machine Learning*, 119:1597–1607, 2020.
- [35] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 2015. doi:10.48550/arXiv.1412.6980.
- [36] Shen Han, Haitao Fu, Yuyang Wu, Ganglan Zhao, Zhenyu Song, Feng Huang, Zhongfei Zhang, Shichao Liu, and Wen Zhang. Himgnn: a novel hierarchical molecular graph representation learning framework for property prediction. *Briefings in Bioinformatics*, 24(5), 2023. doi:10.1093/bib/bbad305.
- [37] Xiaomin Fang, Lihang Liu, Jieqiong Lei, Donglong He, Shanzhuo Zhang, Jingbo Zhou, Fan Wang, Hua Wu, and Haifeng Wang. Geometry-enhanced molecular representation learning for property prediction. *Nature Machine Intelligence*, 4(2):127–134, 2022.

- [38] 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. 2023. doi:10.26434/chemrxiv-2022-jjm0j-v4.
- [39] Xiangxiang Zeng, Hongxin Xiang, Linhui Yu, Jianmin Wang, Kenli Li, Ruth Nussinov, and Feixiong Cheng. Accurate prediction of molecular properties and drug targets using a self-supervised image representation learning framework. *Nature Machine Intelligence*, 4(11):1004–1016, 2022.
- [40] Tom Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.
- [41] Tom Chai and Ronald R. Draxler. Root mean square error (rmse) or mean absolute error (mae)?—arguments against avoiding rmse in the literature. *Geoscientific Model Development*, 7:1247–1250, 2014.
- [42] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002.
- [43] W. M. Pardridge. The blood-brain barrier: Bottleneck in brain drug development. *NeuroRx*, 2(1):3–14, 2005.
- [44] C. A. Lipinski. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 46(1-3):3–26, 2001.
- [45] P. Atkins and J. de Paula. Physical chemistry. 2017.
- [46] P. Ertl, B. Rohde, and P. Selzer. Fast calculation of molecular polar surface area as a sum of fragment-based contributions. *Journal of Medicinal Chemistry*, 43(20):3714–3717, 2000.
- [47] S. Ekins, J. Mestres, and B. Testa. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *British Journal of Pharmacology*, 152(1):9–20, 2002.
- [48] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. Neural message passing for quantum chemistry. *Proceedings of the 34th International Conference on Machine Learning*, 70:1263–1272, 2017.
- [49] Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard Zemel. Gated graph sequence neural networks. In *International Conference on Learning Representations*, 2016.
- [50] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, and Others. Pytorch: An imperative style, high-performance deep learning library. <https://pytorch.org>, 2019.
- [51] Matthias Fey and Jan E. Lenssen. Fast graph representation learning with pytorch geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.
- [52] Greg Landrum. Rdkit: Open-source cheminformatics. <https://www.rdkit.org>, 2013.

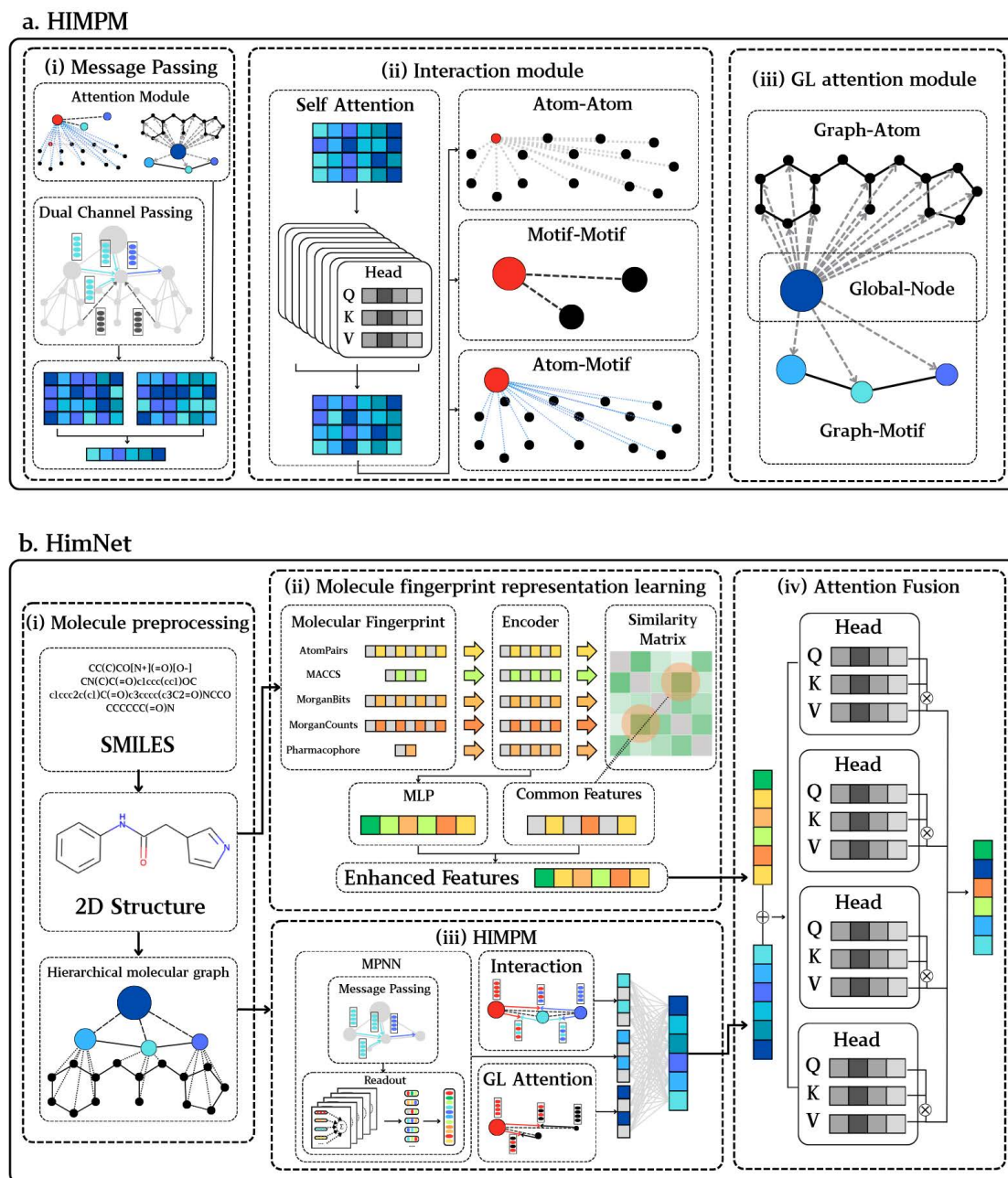


Figure 1: Overview of the proposed HimNet framework. (a) Hierarchical Interaction Message Passing Module (HIMPM). (i) Message Passing: Combines a dual-channel architecture consisting of directed message passing (D-MPNN) and a hierarchical interactive attention mechanism. (ii) Interaction Module: Models multi-scale interactions including atom-atom, motif-motif, and atom-motif relationships using cross-hierarchical attention mechanisms. (iii) Global-Local Attention Module: Aggregates information from atom-level and motif-level nodes into a unified graph-level representation through hierarchical attention fusion. (b) HimNet Architecture. (i) Molecule Preprocessing: Converts SMILES strings into 2D molecular structures and constructs hierarchical molecular graphs comprising atom, motif, and global nodes. (ii) Molecular Fingerprint Representation Learning: Encodes multiple types of molecular fingerprints and extracts common features via similarity-guided fusion strategies. (iii) HIMPM: Applies the hierarchical message passing and interaction mechanisms described in (a) to the hierarchical graph representations. (iv) Attention Fusion: Integrates multi-source features from fingerprints and hierarchical graphs using multi-head self-attention to produce expressive molecular embeddings for downstream tasks.

| Dataset | BACE | BBBP | Tox21 | SIDER | ClinTox | ESOL | Freesolv | Lipophilicity |
|-----------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| MolCLR | 0.890(0.003) | 0.736(0.005) | 0.798(0.007) | 0.680(0.011) | 0.932(0.017) | 1.110(0.010) | 2.200(0.200) | 0.650(0.080) |
| GEM | 0.856(0.011) | 0.724(0.004) | 0.781(0.001) | 0.672(0.004) | 0.901(0.013) | 0.798(0.029) | 1.877(0.094) | 0.660(0.008) |
| ImageMol | 0.839(0.005) | 0.739(0.002) | 0.773(0.001) | 0.660(0.001) | 0.851(0.014) | 0.970(0.070) | 2.020(0.070) | 0.720(0.010) |
| MolMCL _{GIN} | 0.850(0.011) | 0.741(0.006) | 0.775(0.003) | 0.667(0.008) | 0.957(0.012) | — | — | — |
| MolMCL _{GPS} | 0.861(0.013) | 0.736(0.007) | 0.790(0.006) | 0.687(0.002) | 0.951(0.005) | — | — | — |
| Uni-Mol | 0.857(0.002) | 0.729(0.006) | 0.796(0.005) | 0.659(0.013) | 0.919(0.018) | 0.788(0.029) | 1.480(0.048) | 0.603(0.010) |
| MESPool | 0.855(0.039) | 0.848(0.046) | 0.787(0.019) | 0.576(0.026) | 0.902(0.065) | 1.276(0.246) | 2.779(0.762) | 0.708(0.042) |
| HimGNN | 0.856(0.034) | 0.928(0.027) | 0.807(0.017) | 0.642(0.023) | 0.917(0.030) | 0.870(0.154) | 1.921(0.474) | 0.632(0.016) |
| HiMol | 0.846(0.002) | 0.732(0.008) | 0.762(0.003) | 0.625(0.003) | 0.808(0.014) | 0.833 | 2.283 | 0.708 |
| FH-GNN ^a | 0.882(0.029) | 0.949(0.016) | 0.824(0.014) | 0.639(0.002) | 0.945(0.046) | 0.904(0.070) | 1.873(0.263) | 0.744(0.091) |
| HimNet | 0.890(0.026) | 0.954(0.020) | 0.826(0.018) | 0.651(0.009) | 0.950(0.051) | 0.710(0.016) | 1.441(0.223) | 0.698(0.037) |

Table 1: Comparative performance of HimNet and baseline models on eight benchmark datasets. For all baselines except FH-GNN^a, results are cited from the original publications. As the original FH-GNN did not disclose its dataset splitting protocol, we reproduced FH-GNN^a under the same scaffold-based split and experimental settings as HimNet to ensure a fair and direct comparison. Results marked in **bold** indicate the best performance.

| Dataset | MetStab | Malaria | LMC-H | LMC-R | LMC-M | LMC _{mean} |
|---------------|---------------------|---------------------|-------------------|-------------------|--------------------|---------------------|
| MPNN | 0.823(0.020) | 1.056(0.037) | 104.0(5.8) | 162.9(27.4) | 125.9(20.6) | 130.9(4.1) |
| GCN | 0.696(0.055) | 1.057(0.043) | 103.7(5.7) | 161.6(26.2) | 126.4(20.2) | 130.6(3.6) |
| GAT | 0.767(0.031) | 1.048(0.040) | 102.6(5.5) | 159.4(24.6) | 124.8(17.3) | 128.9(3.6) |
| GIN | 0.773(0.041) | 1.045(0.038) | 102.7(6.3) | 158.4(22.9) | 126.7(19.4) | 129.3(2.2) |
| FH-GNN | 0.876(0.014) | 1.053(0.035) | 96.664(9.6) | 144.785(8.2) | 167.860(25.9) | 136.436(6.7) |
| MolCLR | 0.746(0.052) | 1.086(0.072) | 98.471(1.4) | 127.396(15.9) | 197.507(41.3) | 141.125(10.2) |
| HiMol | 0.844(0.034) | 1.081(0.056) | 95.785(16.8) | 112.170(11.4) | 143.301(36.6) | 117.085(10.1) |
| HimGNN | 0.699(0.115) | 1.084(0.113) | — | — | — | — |
| HimNet | 0.896(0.008) | 1.029(0.030) | 91.7(11.9) | 109.3(0.7) | 132.1(39.5) | 111.0(9.8) |

Table 2: Comparative performance of HimNet and four widely used GNN baselines (MPNN, GCN, GAT, GIN) and four hierarchical GNN baselines(FHGNN,MolCLR,HiMol,HimGNN) on MetStab, Malaria, and LMC (human, rat, mouse) tasks. All experiments were conducted under identical scaffold-based data splits. For each model, results are averaged over multiple runs with three different splits and random initializations; mean and standard deviation are reported. Results marked in **bold** indicate the best performance.

| Variant | BACE | BBBP | Tox21 | SIDER | ClinTox | ESOL | Freesolv | Lipophilicity |
|---------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| HimNet | 0.876(0.002) | 0.962(0.002) | 0.835(0.006) | 0.651(0.002) | 0.983(0.002) | 0.712(0.017) | 1.226(0.033) | 0.711(0.002) |
| w/o M | 0.868(0.004) | 0.959(0.005) | 0.825(0.004) | 0.645(0.005) | 0.982(0.007) | 0.738(0.008) | 1.242(0.020) | 0.723(0.005) |
| w/o I | 0.863(0.006) | 0.958(0.002) | 0.827(0.004) | 0.636(0.004) | 0.979(0.002) | 0.768(0.012) | 1.310(0.023) | 0.728(0.007) |
| w/o C | 0.864(0.001) | 0.960(0.004) | 0.831(0.008) | 0.631(0.008) | 0.975(0.009) | 0.826(0.010) | 1.656(0.056) | 0.764(0.013) |
| w/o A | 0.865(0.002) | 0.952(0.004) | 0.828(0.001) | 0.647(0.002) | 0.971(0.003) | 0.783(0.019) | 1.273(0.059) | 0.717(0.002) |
| w/o H | 0.865(0.006) | 0.955(0.009) | 0.826(0.003) | 0.645(0.010) | 0.972(0.012) | 0.792(0.015) | 1.238(0.035) | 0.731(0.008) |

Table 3: Ablation results for HimNet: performance of the full model and its variants with each core module removed. The best result for each dataset is highlighted in bold. For clarity, the abbreviations for the ablated variants are defined as follows. **w/o M**: removes the Dual message passing module; **w/o I**: removes the Hierarchical interaction module; **w/o C**: removes the Consensus fingerprint enhancement module; **w/o A**: removes the Attention-guided fusion module; **w/o H**: removes the Hierarchical message passing module. Results marked in **bold** indicate the best performance.

| Method | ClinTox | | | | Freesolv | | | |
|---------------|---------------------|---------------|----------------|-----------------|---------------------|---------------|----------------|-----------------|
| | AUC | Params | Training Time | Inference Speed | RMSE | Params | Training Time | Inference Speed |
| HiMoI | 0.808(0.014) | 54.81M | 2.72min | 1.359ms | 2.283 | 54.80M | 0.65min | 2.002ms |
| FH-GNN | 0.945(0.046) | 17.02M | 30.57min | 21.070ms | 1.873(0.263) | 17.02M | 17.23min | 18.881ms |
| HimNet | 0.950(0.051) | 12.94M | 42.35min | 24.127ms | 1.441(0.223) | 12.94M | 7.32min | 17.922ms |

Table 4: Performance comparison with HiMoI and FH-GNN. Params indicates the total number of model parameters, Training Time denotes the total time required for model training on the complete dataset, and Inference Speed represents the average time taken to process a single molecule. Results marked in **bold** indicate the best performance.

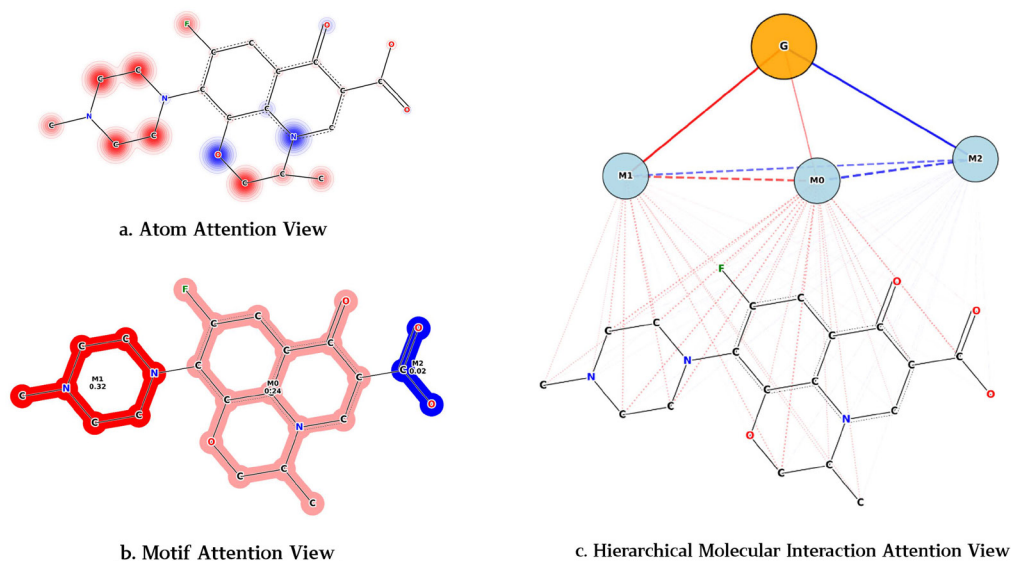


Figure 2: Hierarchical attention visualization for molecule CC1COc2c(N3CCN(C)CC3)c(F)cc3c(=O)c(C(=O)O)cn1c23. **(a) Atomic-level attention:** Highlights atom-wise contributions to BBB permeability, with red indicating positive and blue negative effects, consistent with chemical intuition that hydrophobic atoms promote and polar atoms inhibit penetration. **(b) Motif-level attention:** Aggregates attention to chemically meaningful substructures, showing enhanced positive attention on aromatic and fluorinated motifs, and negative attention on polar or acidic fragments. **(c) Hierarchical interaction attention:** Integrates atomic, motif, and global graph levels, revealing context-dependent interactions—such as competitive and cooperative effects—between motifs and their surrounding atomic environments, thereby supporting chemically interpretable property prediction.

| Hyperparameter | Value |
|-------------------------|--|
| Batch size | 64 |
| Learning rate | 0.0001 |
| Hidden layer dimension | 512 |
| Molecular encoder depth | 7 layers |
| Cross-attention heads | 8 |
| Fusion attention heads | 4 |
| Dropout rate | 0.1 |
| Training epochs | 100 |
| Optimizer | Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$) |

Table 5: Hyperparameter settings for the HimNet model

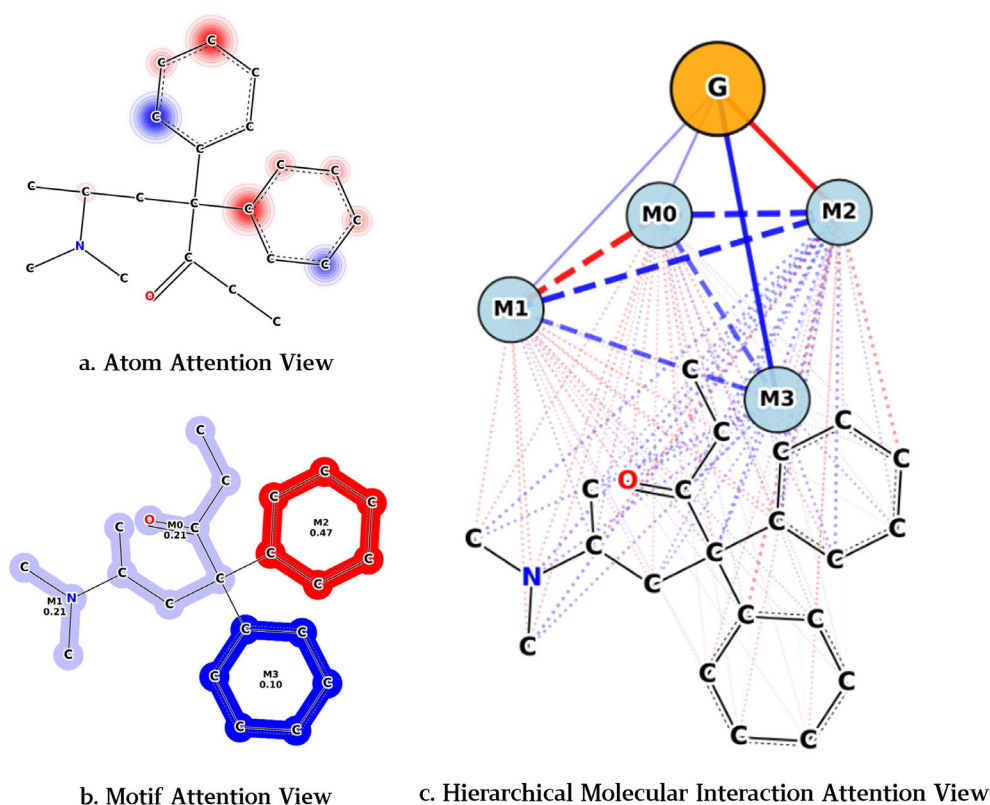


Figure 3: Hierarchical attention visualization of CCC(=O)C(CC(C)N(C)C)(c1ccccc1)(c1ccccc1), illustrating chemically interpretable multi-level feature interactions for BBB permeability prediction. **(a) Atomic-level attention:** The model assigns strong positive attention to the carbon atoms of both phenyl rings and the adjacent aliphatic carbons, reflecting their role in enhancing hydrophobicity and promoting BBB permeability. In contrast, negative attention is focused on the carbonyl oxygen and the nitrogen atom of the tertiary amine side chain, indicating these polar centers hinder BBB penetration. **(b) Motif-level attention:** The two phenyl rings, though structurally similar, are distinguished by the model: one is assigned strong positive attention, while the other receives negative attention, reflecting their distinct chemical environments and contextual influence from neighboring groups. The carbonyl and amine motifs both receive negative attention, consistent with their polar and hydrogen-bonding character. **(c) Hierarchical interaction attention:** The integrated view demonstrates that the model captures competitive and complementary relationships among motifs. The strongly positive aromatic motif (M2) competes with the negatively weighted aromatic motif (M3) and the polar carbonyl and amine motifs (M0, M1) in determining the global property. Notably, the context-dependent differentiation of the two phenyl motifs highlights the model’s ability to capture non-additive and position-sensitive effects, aligning with chemical intuition regarding the interplay between hydrophobic and polar regions in BBB permeability.

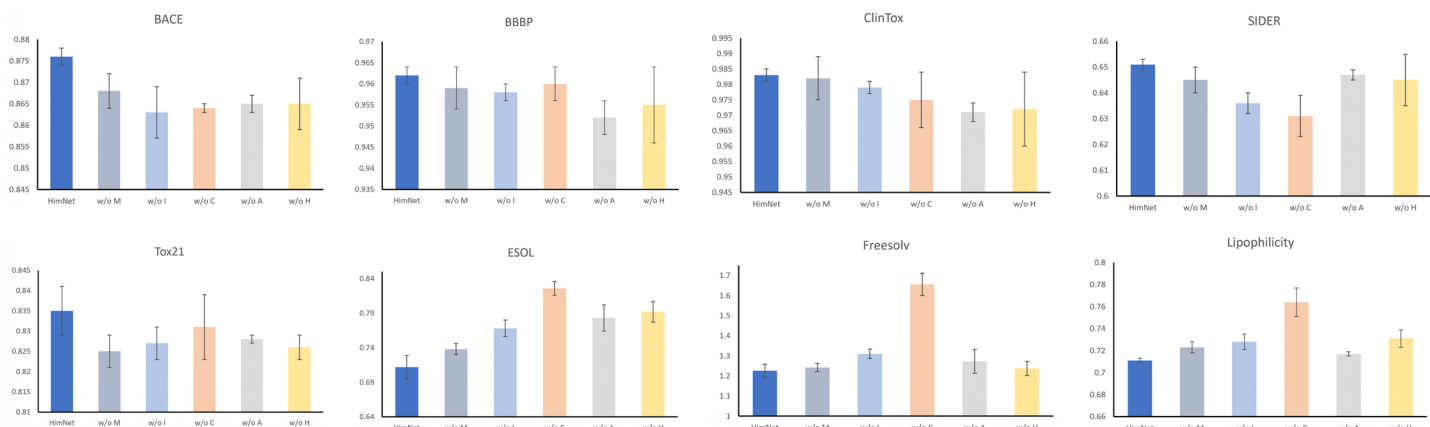


Figure 4: Performance comparison of HimNet and its ablated variants on eight benchmark datasets. Each bar shows the mean performance metric of the full model (HimNet) or a variant with a specific module removed (see legend for abbreviations). Error bars represent the standard deviation across repeated runs. Removing any module leads to a consistent performance drop, confirming the necessity of each component.

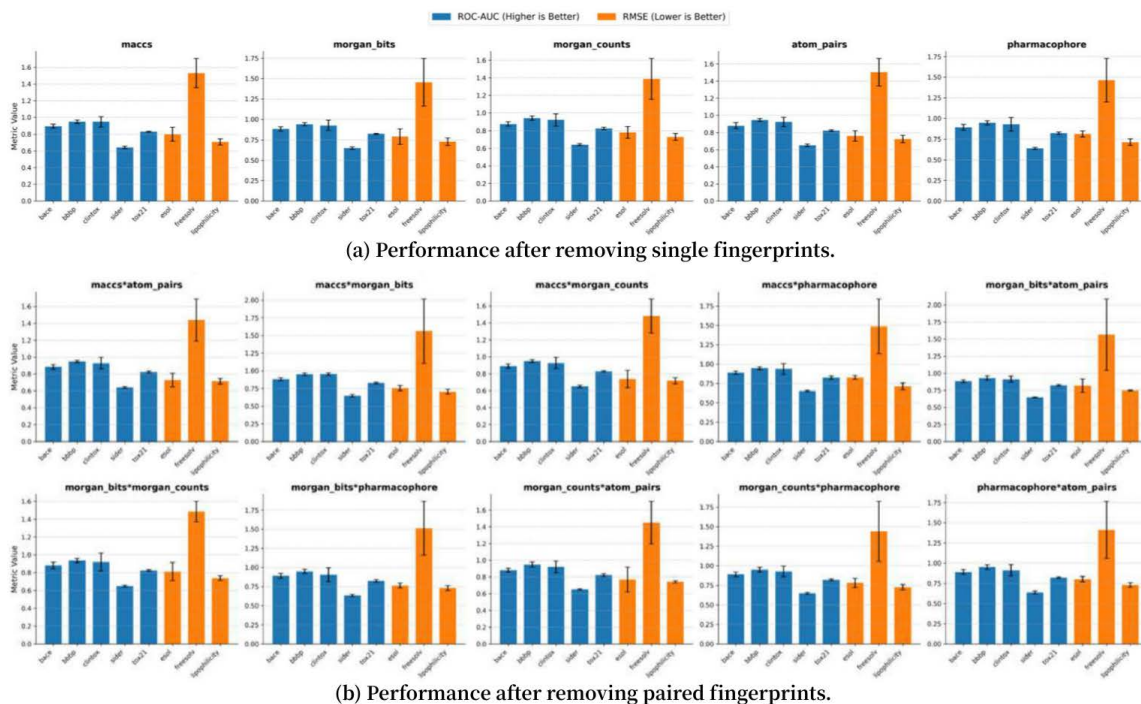


Figure 5: Visual analysis of the fingerprint ablation study. (a) Performance after removing single fingerprints: Each subplot corresponds to a model variant where a specific single fingerprint has been ablated. The results highlight the task-dependent importance of each individual fingerprint, as shown by the varying degrees of performance degradation (lower ROC-AUC or higher RMSE) across different tasks. (b) Performance after removing paired fingerprints: These subplots illustrate the model's performance when two fingerprint types are simultaneously ablated. The generally more substantial performance degradation compared to the single-fingerprint ablations provides strong evidence for the strong synergistic effects among the different fingerprint types.

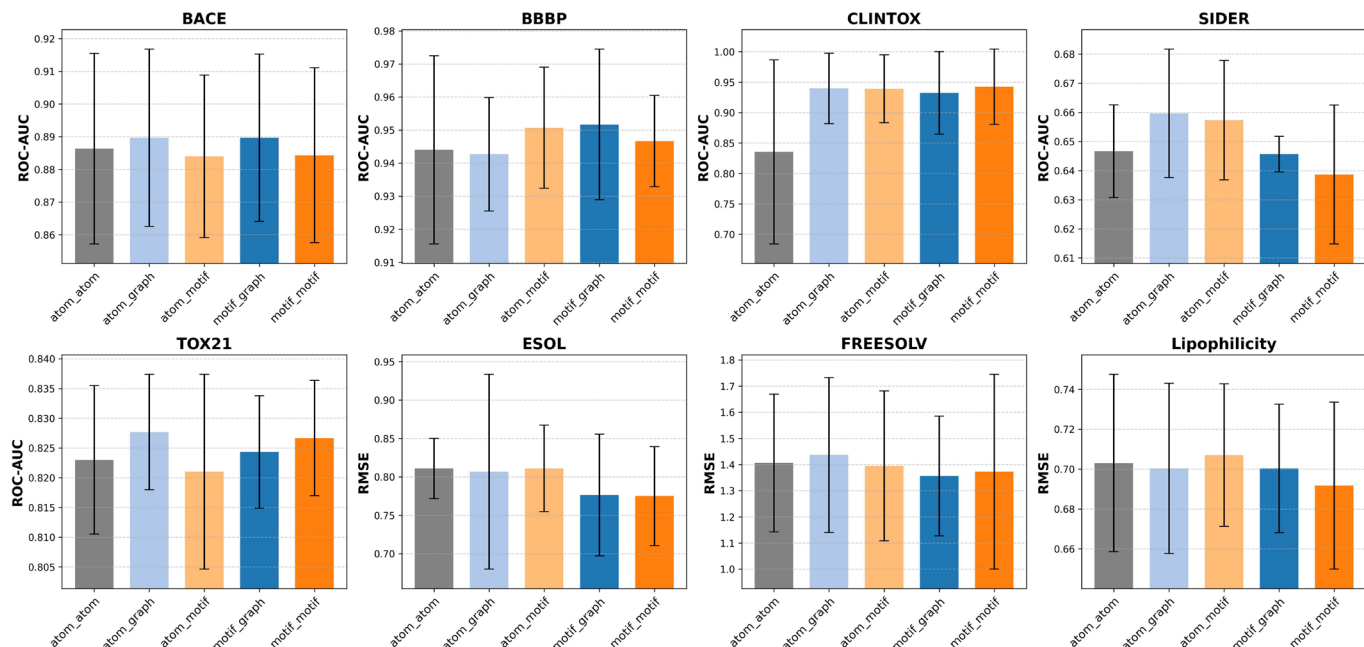


Figure 6: Visual analysis of the ablation study on hierarchical interaction mechanisms. Each subplot corresponds to a specific dataset, and the bars represent the model’s performance after the interaction labeled on the x-axis (e.g., *atom_atom*) has been removed. Performance interpretation is task-dependent: for classification tasks (BACE, BBBP, CLINTOX, SIDER, TOX21), the metric is ROC-AUC, where a lower bar signifies a greater performance drop. Conversely, for regression tasks (ESOL, FREESOLV, Lipophilicity), the metric is RMSE, where a higher bar indicates a more significant degradation. Error bars denote the standard deviation over three independent runs. The results compellingly demonstrate that the contribution of each interaction is highly task-dependent, validating our model’s multifaceted design.

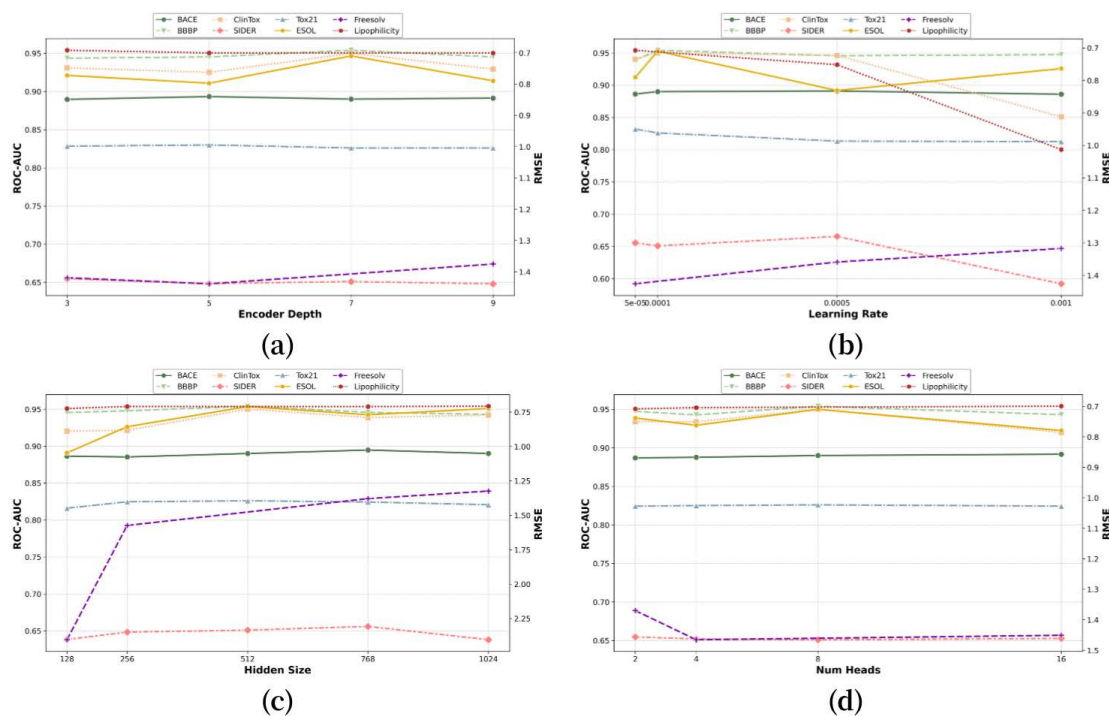


Figure 7: Sensitivity analysis of HimNet with respect to four key hyperparameters. The subplots illustrate the model’s performance variation in response to changes in: (a) encoder depth, (b) learning rate, (c) hidden size, and (d) number of attention heads. For each analysis, one parameter is varied while the others are held constant at their optimal values. The results demonstrate the model’s overall robustness and validate our final hyperparameter selections.

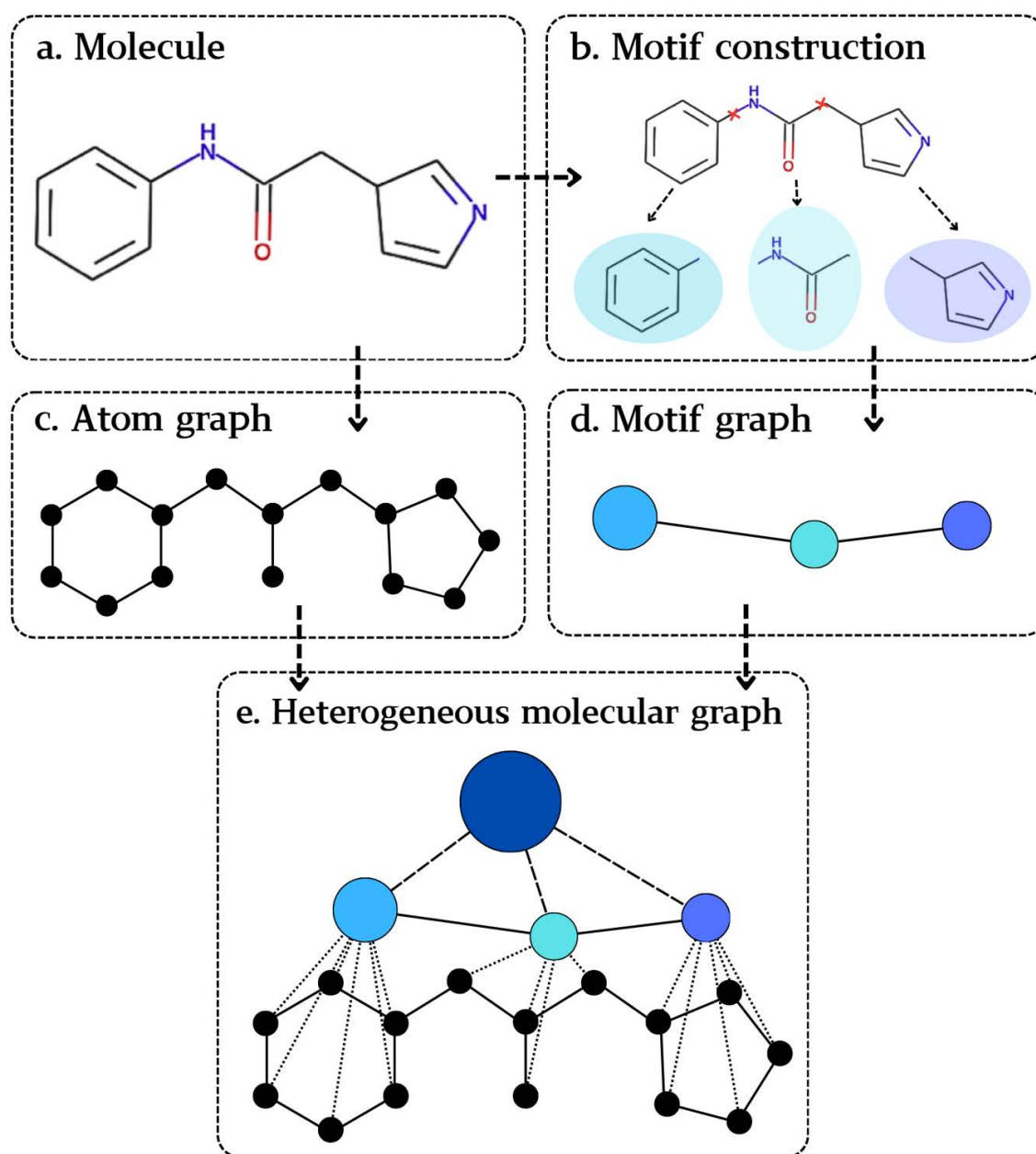


Figure 8: Schematic of the hierarchical heterogeneous molecular graph construction process. (a) The original 2D molecular structure. (b) The molecule is decomposed into chemical motifs using the BRICS algorithm. (c) The corresponding atom-level graph, where nodes are atoms and edges are chemical bonds. (d) The corresponding motif-level graph, where nodes are motifs and edges represent covalent connections between them. (e) The final hierarchical heterogeneous molecular graph, which integrates the atom and motif graphs and includes a central global node connected to all motifs to learn a holistic molecular representation.

| Data class | Dataset | # molecules | # tasks | Metric | Type |
|-----------------|--|-------------|---------|--------|----------------|
| Physicochemical | ESOL (estimating aqueous solubility) | 1 128 | 1 | RMSE | Regression |
| | FreeSolv (hydration free energy) | 642 | 1 | RMSE | Regression |
| | Lipophilicity (octanol–water partition coefficient LogP) | 4 200 | 1 | RMSE | Regression |
| Bioactivity | BACE (β -secretase 1 inhibitor activity) | 1 513 | 1 | AUC | Classification |
| | Malaria (anti-malarial EC ₅₀) | 9 998 | 1 | RMSE | Regression |
| Toxicity | Tox21 (12 toxicity endpoints) | 7 831 | 12 | AUC | Classification |
| | SIDER (27 adverse drug reaction labels) | 1 427 | 27 | AUC | Classification |
| | ClinTox (clinical trial toxicity) | 1 478 | 2 | AUC | Classification |
| Pharmacokinetic | BBBP (blood–brain barrier penetration) | 2 039 | 1 | AUC | Classification |
| | LMC (liver microsomal clearance in human, rat, mouse) | 8 755 | 3 | RMSE | Regression |
| | MetStab (metabolic stability half-life) | 2 267 | 1 | AUC | Classification |

Table 6: Comprehensive summary of the 11 benchmark datasets evaluated in this work, encompassing physicochemical, bioactivity, toxicity, and pharmacokinetic properties. The table details the number of molecules, tasks, metrics, and task types for each dataset, including both widely used MoleculeNet benchmarks and several challenging, less-explored datasets, thereby ensuring a thorough assessment of model performance and generalization.