BIG DATA MINING AND ANALYTICS
ISSN 2096-0654 20/25 pp858–888
DOI: 10.26599/BDMA.2024.9020028
Volume 7, Number 3, September 2024

Impact of Domain Knowledge and Multi-Modality on Intelligent Molecular Property Prediction: A Systematic Survey

Taojie Kuang, Pengfei Liu, and Zhixiang Ren*

Abstract: The precise prediction of molecular properties is essential for advancements in drug development, particularly in virtual screening and compound optimization. The recent introduction of numerous deep learningbased methods has shown remarkable potential in enhancing Molecular Property Prediction (MPP), especially improving accuracy and insights into molecular structures. Yet, two critical questions arise: does the integration of domain knowledge augment the accuracy of molecular property prediction and does employing multi-modal data fusion yield more precise results than unique data source methods? To explore these matters, we comprehensively review and quantitatively analyze recent deep learning methods based on various benchmarks. We discover that integrating molecular information significantly improves Molecular Property Prediction (MPP) for both regression and classification tasks. Specifically, regression improvements, measured by reductions in Root Mean Square Error (RMSE), are up to 4.0%, while classification enhancements, measured by the area under the receiver operating characteristic curve (ROC-AUC), are up to 1.7%. Additionally, we discover that, as measured by ROC-AUC, augmenting 2D graphs with 3D information improves performance for classification tasks by up to 13.2% and enriching 2D graphs with 1D SMILES boosts multi-modal learning performance for regression tasks by up to 9.1%. The two consolidated insights offer crucial guidance for future advancements in drug discovery.

Key words: Molecular Property Prediction (MPP); Deep Learning (DL); domain knowledge; multi-modality; drug discovery

1 Introduction

The field of drug development has always been at the forefront of adopting innovative scientific techniques to enhance the discovery and optimization of therapeutic compounds. Central to this process is the prediction of molecular properties, a task that bears significant implications for drug screening and compound optimization[1]. Accurately predicting key molecular properties can significantly reduce the time and resources required in drug development, thereby hastening the journey towards innovative medical treatments.

In the landscape of computational methods for Molecular Property Prediction (MPP), Deep Learning

• Taojie Kuang is with Peng Cheng National Laboratory, Shenzhen 518000, China, and also with School of Future Technology, South China University of Technology, Guangzhou 511442, China. E-mail: kuangtj@pcl.ac.cn.
• Pengfei Liu is with Peng Cheng National Laboratory, Shenzhen 518000, China, and also with School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, China. E-mail: liupf@pcl.ac.cn.
• Zhixiang Ren is with Peng Cheng National Laboratory, Shenzhen 518000, China. E-mail: jason.zhixiang.ren@outlook.com.
* To whom correspondence should be addressed.
Manuscript received: 2024-02-18; revised: 2024-04-10; accepted: 2024-04-15

(DL) has recently emerged as a transformative force, distinguishing itself markedly from traditional techniques, such as Quantitative Structure-Activity Relationships (QSAR) and molecular dynamics simulations. While conventional methods have laid the groundwork, DL significantly advances accuracy and analysis depth, enabling a more intricate exploration of the relationships between molecular structures and their properties[2].

Despite the advancements in DL for MPP, the field continues to face ongoing evolution and challenges. Two significant trends are currently shaping the field. The first trend is the increasing integration of domain knowledge into DL models. This includes a broad spectrum of scientific information, such as chemical and physical property relations, atom and bond characteristics, and detailed insights into functional groups and molecular fragments. The integration of this knowledge aims to enhance the predictive accuracy of these models. This leads us to the critical question: Does more comprehensive domain knowledge actually improve the effectiveness of MPP? The second trend is the rising adoption of multi-modality techniques, which involve the fusion of various data types, like sequence-based, graph-based, and pixel-based formats. This approach is driven by the goal of achieving more accurate predictions in a field characterized by its complexity and data diversity, prompting the question: Is multi-modality more effective for MPP than methodologies that rely on uni-model data source? To explore these questions, our paper begins with an in-depth review of the current DL approaches in MPP, focusing on how the domain knowledge and multi-modal data integrate on encoder architecture and training strategy.

Our review begins with an examination of various unique data encoder architectures for MPP, such as Recurrent Neural Network (RNN)[3–5], Graph Neural Network (GNN)[6–10], Transformer[11–14], and Convolutional Neural Network (CNN)[15–17] models, and also reviews the multi-modal methods[18–21]. We focus on how these architectures are aligned with existing molecular structural knowledge and their integration of domain knowledge. This exploration highlights the synergy between advanced computational techniques and fundamental molecular understanding, a crucial aspect in enhancing the accuracy of MPP. Also, we review a variety of training strategies, such as self-supervised[22–24], semi-supervised[25–27], transfer learning[28–30], and multi-tasks learning[31, 32]. A particular emphasis is placed on strategies that effectively utilize unlabeled data, a vital consideration given the frequent scarcity of labeled data in this domain, and we focus on how the domain knowledge and multi-modal data to be used in the training strategy. Accompanying this review are comprehensive diagrams that systematically elucidate the nuances of these encoder architectures and training strategies, offering a clearer understanding of their complex mechanisms. The overview of our paper is shown in Fig. 1.

Our study then proceeds to empirically evaluate these DL methods, utilizing pivotal benchmarks like MoleculeNet[33]. These benchmarks, encompassing a diverse range of datasets each focused on specific molecular properties, allow for an extensive assessment of different DL approaches. A key aspect of our analysis is determining the impact of multi-modality techniques versus single modeling. Specifically, we investigate the effectiveness of integrating atom-bond level domain knowledge and substructures, such as functional groups and fragments, into the models. Additionally, we quantify the contributions of different data formats and conduct experiments to ascertain whether multi-modal fusion can enhance the generalization performance of the models. This evaluation not only provides comparative insights into the varied methods, but also seeks to pinpoint essential factors that bolster the efficacy of DL in MPP.

In summary, our main contributions are as follows:

- We identify two pivotal issues when applying DL for MPP: domain knowledge integration and multi-modal data utilization.
- We comprehensively review DL methods for MPP, featuring in-depth analyses of encoder architectures and training strategies.
- We discover that integrating molecular substructure information results in a 4.0% improvement on average in regression tasks and a 1.7% increase on average in classification tasks.
- We discover that enriching 2D graph models with 1D SMILES or 3D information boosts multi-modal learning, enhancing performance by 9.1% to 13.2% over single-modality models.

2 Molecular Modality

In the field of molecular science, a wide variety of

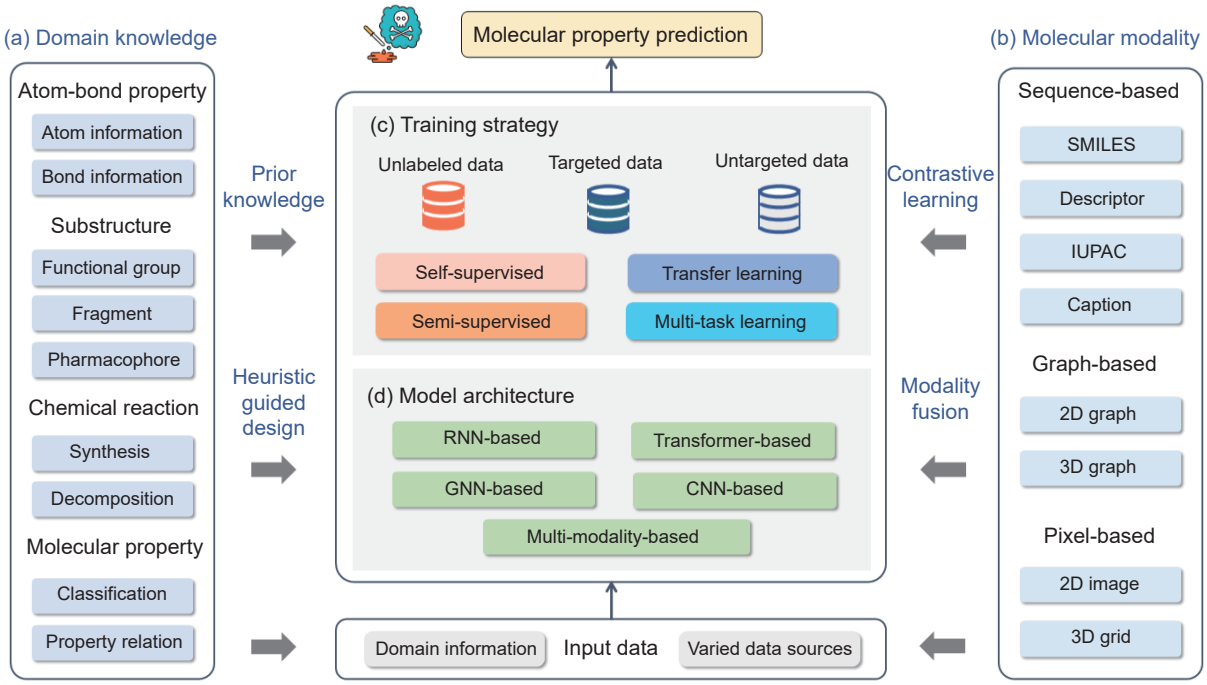

Fig. 1 Overview of our survey. we review the impact of domain knowledge and multi-modality on molecular property prediction from three critical aspects: input data, model architectures, and training strategy. The detailed information is explained in the following sections.

molecular modalities have been developed, which are essential for computational modeling and analysis. These formats are generally classified into three main types: text-based, graph-based, and pixel-based formats. Each type offers unique insights into molecular structures, contributing significantly to various aspects of molecular analysis. These diverse formats are illustrated in Fig. 2, which showcases the array of molecular modality available for DL methods.

2.1 Sequence-based data

Text-based formats are among the most commonly used representations in MPP due to their simplicity and efficiency. The most prominent of these is Simplified Molecular Input Line Entry System (SMILES)[34], which encodes molecules in linear strings, representing atoms and bonds in a compact and readable format. Variants of SMILES, such as Canonical SMILES[35] and Isomeric SMILES[36], offer additional specificity, including stereochemistry information. Other notable text-based formats include molecular fingerprints, like ECFP[37], Morgan, and MACCS[38], which encode the presence of certain molecular features, and self-referencing embedded strings (namely SELFIES)[39], a newer format designed for robustness in machine learning applications. Additionally, IUPAC[40] and InChI[41] codes are vital text-based molecular representations. IUPAC provides systematic chemical nomenclature for clear scientific communication, while InChI offers standardized textual identifiers for chemical substances. These formats facilitate various computational tasks, from database searching to the generation of novel molecules using AI.

2.2 Graph-based data

In drug discovery, graph-based representations, which depict atoms as nodes and bonds as edges, effectively capture molecular structures, making them ideal for analyzing both topological and relational aspects of molecules. The method includes the use of a 2D adjacency matrix or a set of edges to outline atom connectivity. This representation can be enhanced with 3D information, such as bond lengths and atom positions, transforming it into a 3D graph. Incorporating a 3D atom distance matrix further enriches this model, offering a comprehensive view of the molecular spatial structure. Graph-based formats, including 2D and 3D molecular structures, are crucial in drug discovery for conducting detailed molecular analyzes and enhancing the understanding of complex

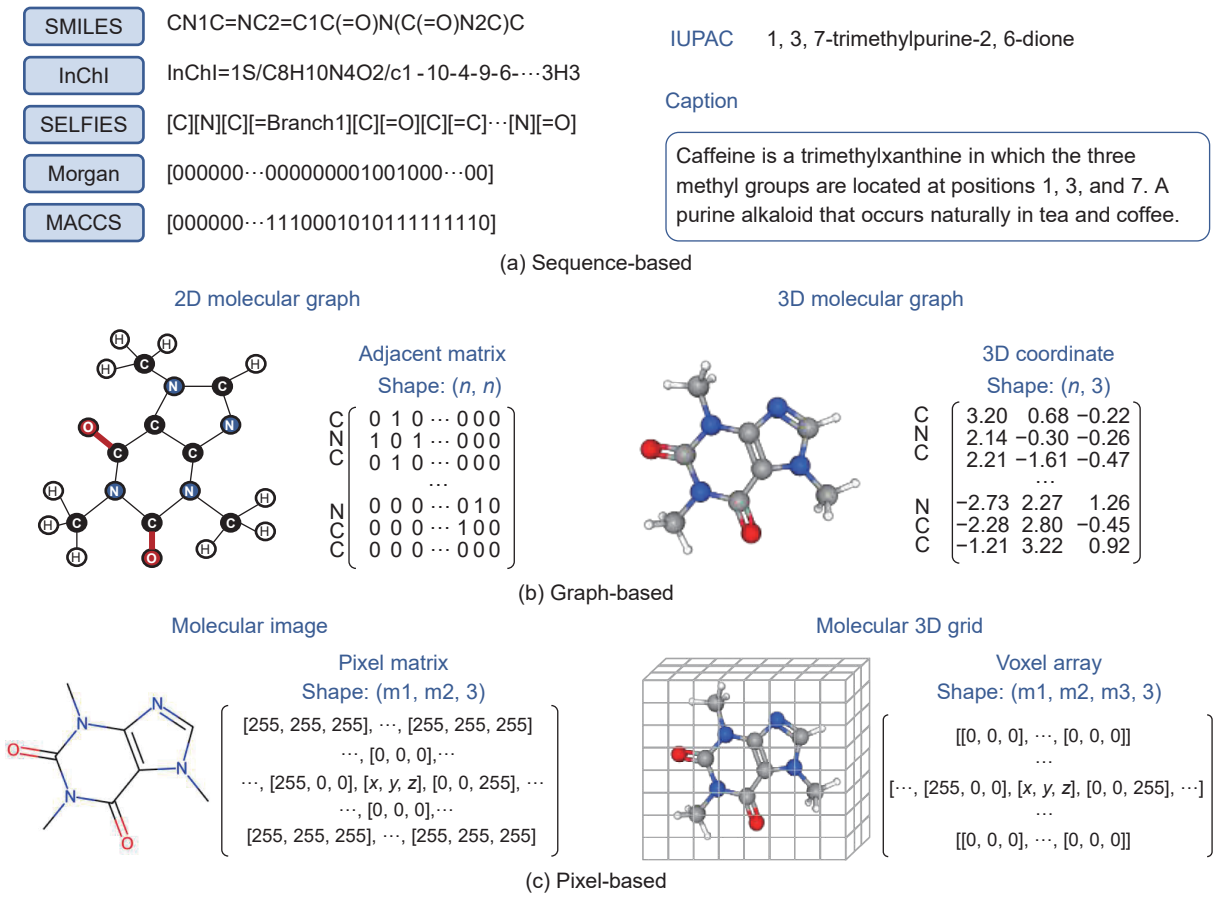

Fig. 2 Molecular modality. Using the caffeine molecule as an example, we demonstrate the various molecular modalities that are essential for MPP. This is demonstrated across three primary categories: sequence-based, graph-based, and pixel-based formats. Each format is derived from the SMILES representation of caffeine, using Python packages such as RDKit and software tools like PyMol. (a) Sequence-based data section includes formats like SMILES and its variants (Canonical and Isomeric SMILES), molecular fingerprints (ECFP, Morgan, MACCS), and SELFIES, highlighting their roles in encoding molecular structures. (b) Graph-based data represent caffeine as a graph with atoms as nodes and bonds as edges, enriched with 3D information for detailed structural insights. (c) Pixel-based data showcase 2D images and 3D grids of caffeine, which are crucial for visual analysis and spatial interpretation.

molecular behaviors.

2.3 Pixel-based data

Pixel-based molecular data formats, such as 2D images and 3D grids, are essential components of molecular property prediction. These formats, easily generated by tools, like RDKit[42] and PyMol[43] for 2D images and Libmolgrid[44] for 3D grids, offer clear and comprehensible visual representations of molecular structures. This visual aspect allows for straightforward human interpretation, aiding in the recognition of molecular patterns and the understanding of spatial relationships in computational modeling. This clarity in visualization is crucial for effectively analyzing molecular geometries and interactions.

3 Domain Knowledge

Domain knowledge, from fields such as physics, chemistry, and biology, plays a crucial role for MPP. This knowledge is methodically grouped into four key categories: atom-bond property, molecular substructure, chemical reactions, and molecular characteristics. Each category is integral for a comprehensive understanding and accurate interpretation of molecular data. Figure 3 showcases these categories in detail, providing an in-depth look at the essential aspects of molecular information interpretation.

3.1 Atom-bond property

In MPP, a deep understanding of atomic and bonding

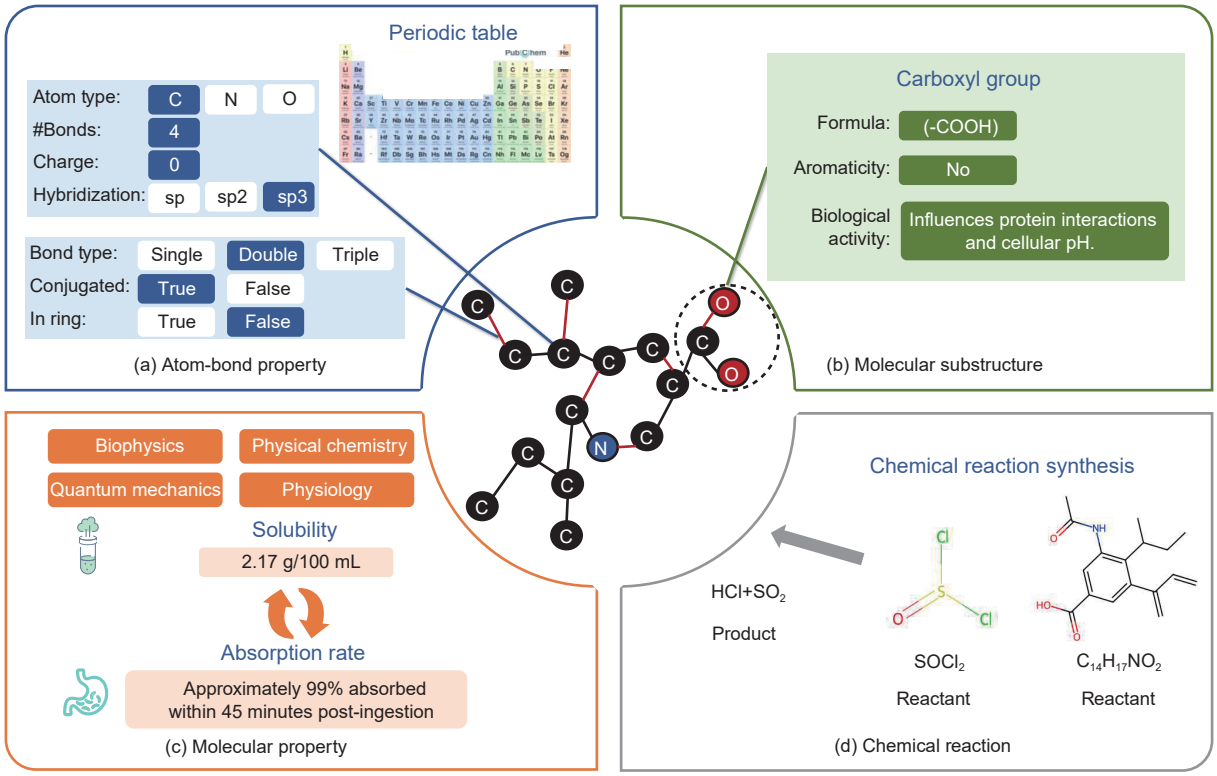

Fig. 3 Molecular domain knowledge. This figure categorizes molecular expert knowledge essential for MPP into four domains, using the molecule C = (CC(= C)c1cc(C(= O)O)cnc1C(C)CC as an example. (a) In the atom-bond property section, we examine aspects, such as the molecule's atomic number, mass, valence, and bond types. (b) Molecular substructure includes the functional groups, molecular fragments, and pharmacophores of this molecule, illustrating their influence on the chemical behavior and interactions. (c) Molecular property covers a range of properties from quantum mechanics to physiology, showcasing how these properties affect the molecule's behavior in drug development. (d) Chemical reaction discusses the mechanisms of molecular transformations, highlighting the molecule's reactivity.

attributes is vital for accurately modeling molecular behaviors. Understanding atomic properties is essential for molecular analysis. For example, isotope numbers influence molecular weight and stability, and chirality is crucial for interactions and reactions within biological systems. Hybridization types impact bonding patterns and molecular geometry. Atomic valence, number, mass, formal charge, and aromaticity significantly influence a molecule's chemistry. Bond attributes, like bond type and stereochemistry, are key in determining molecular connectivity, reactivity, and shape, influencing interactions with biological targets. The direction and length of bonds also provide insights into spatial arrangement. These detailed atomic and bond attributes collectively provide a comprehensive framework for molecular structure analysis, essential for effective predictive modeling in drug discovery.

3.2 Molecular substructure

In the realm of MPP, a deep comprehension of molecular substructures is indispensable. These substructures, including functional groups, molecular fragments, and pharmacophores, are fundamental in dictating the functions and interactions of molecule.

These substructures, such as functional groups, molecular fragments, and pharmacophores, play key roles in understanding a molecule's behavior. Functional groups, such as hydroxyl (—OH) and carboxyl (—COOH), are specific group of atoms within a molecule that is responsible for the characteristic chemical reactions of that molecule, and are particularly influential in determining a molecule's chemical behavior and interactions. For example, a hydroxyl group can significantly increase water solubility, thereby impacting a drug's absorption, distribution, and overall pharmacokinetics.

Molecular fragments are larger portions of molecules, encompassing various structural elements like rings or chains. Similarly, molecular fragments like benzene rings affect a molecule's stability and

electronic, which in turn can alter its interaction with biological receptors or enzymes, impacting biological activity. Common molecular fragment methods are Breaking of Retrosynthetically Interesting Chemical Substructure (BRICS)[45], REtrosynthetic Combinatorial Analysis Procedure (RECAP)[46], Murcko scaffolds[47], eMolFrags[48], and rdScaffoldNetwork[49].

A pharmacophore is an abstract representation of the molecular features that are necessary for a molecule to interact with a specific biological target to produce a desired biological effect. A pharmacophore with both a hydrogen bond donor and an acceptor in a specific spatial arrangement, for instance, can be crucial for binding to biological targets like enzymes or receptors, influencing the molecule's effectiveness as a therapeutic agent. The accurate identification and understanding of these substructures are key to developing new pharmaceuticals, offering detailed insights into molecular interactions.

3.3 Chemical reaction

Chemical reactions involve the transformation of substances through the breaking and forming of chemical bonds, leading to the creation of new molecules with specific properties. For example, in the reaction C = CC(= C)c1cc(C(= O)O)cc(NC(= O)C) c1C(C)CC + SOCl2 → C = CC(= C)c1cc(C(= O)O) cnc1C(C)CC + HCl + SO2, the reactant interacts with thionyl chloride, resulting in a new product plus byproducts. This process highlights the roles of reactants, products, and catalysts in affecting reaction outcomes and mechanisms. Such knowledge is vital for predicting reaction paths, designing new molecules with desired properties, and developing effective pharmaceuticals and novel compounds.

3.4 Molecular property

MPP in drug discovery is a multidisciplinary field, with each discipline offering detailed insights into molecular behavior. Quantum mechanics, for example, delve into electronic properties like ionization potentials, crucial for understanding reaction mechanisms. Physical-chemistry examines molecular stability, reactivity, and phase behaviors, which can significantly impact drug formulation. Biophysics explores molecular interactions within biological systems, crucial for drug-target binding studies. Physiology, on the other hand, assesses drug effects at an organismal level, influencing pharmacodynamics and pharmacokinetics. These interconnected properties, such as how a drug's solubility impacts absorption and bioavailability, highlight the need for a comprehensive understanding across levels, from atomic to organismal, to predict molecular properties accurately and develop effective pharmaceuticals. This integrative approach, encompassing everything from electron distribution to organismal response, is vital in the nuanced field of drug development.

4 Modeling Method

Our paper provides a concise yet comprehensive examination of current DL methods in MPP. We first review molecular encoder architectures, and explore how these encoders align with the prior structural knowledge of molecules and how domain knowledge is integrated into them. Our review further emphasizes the utilization of unlabeled data, encompassing an exploration of self-supervised, semi-supervised, transfer learning, and multi-task learning strategies. To aid in understanding these intricate concepts, our paper includes detailed diagrams, which elucidate these advanced computational methods and their integration with fundamental molecular insights, thereby contributing to the advancement of MPP.

4.1 Encoder

In MPP, encoder architectures play a key role in transforming raw molecular data into meaningful representations. A range of encoder architectures, each suited to a particular molecular modality and complexity, are examined in this section. We categorize four main types of encoders for single data sources: RNN-based, GNN-based, Transformer-based, and CNN-based. Each type is analyzed for its alignment with molecular prior structural knowledge and the integration of domain-specific information. Additionally, we examine multi-modality based encoders, which handle multiple data sources, highlighting their unique characteristics, applications, and the challenges they address in molecular representation learning. The detailed aspects of these encoder architectures are illustrated in the Figs. 4 and 5.

4.1.1 RNN-based method

RNNs, like Long Short-Term Memory (LSTM)[59] and Gated Recurrent Unit (GRU)[60], are adept at processing sequential data, with a unique internal

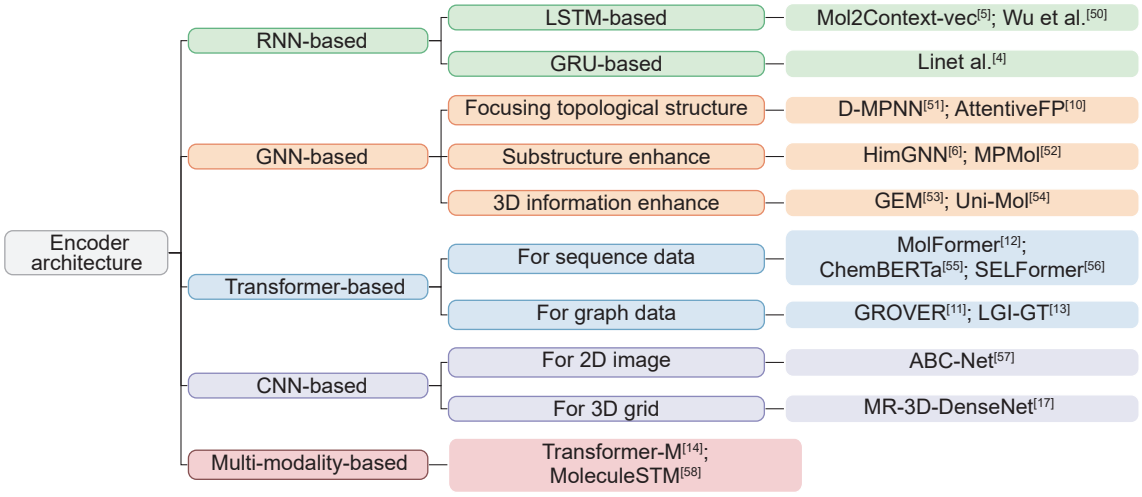

Fig. 4 Molecular encoder method summary. We categorize molecular encoder method into five types: RNN-based, GNN-based, Transformer-based, CNN-based, and multi-modality-based. For each category, key techniques and notable advancements utilized in various influential studies are highlighted, showcasing the evolution and diversification of approaches in molecular encoding.

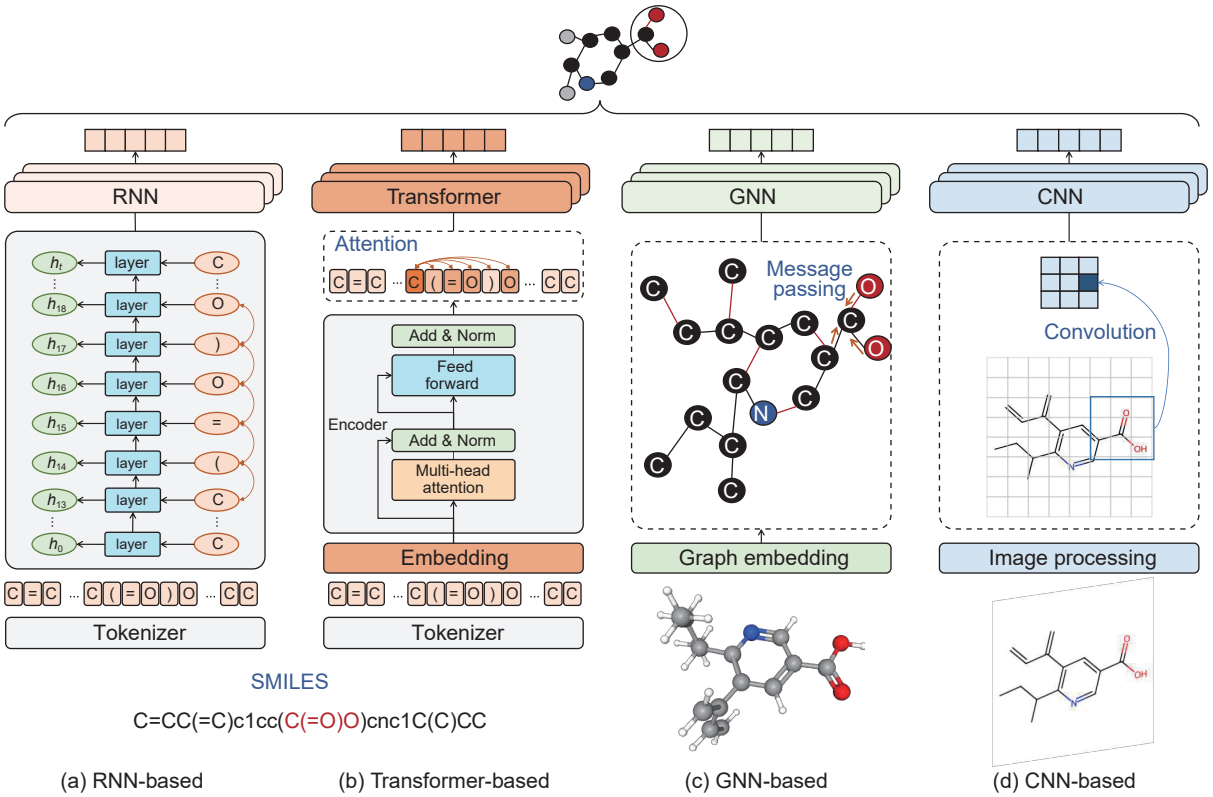

Fig. 5 Molecular encoder architectures. This figure categorizes molecular encoder architectures for single modality into four types: RNN-based, Transformer-based, GNN-based, and CNN-based. Each type is assessed for its ability to capture information about functional groups like carboxyl groups (—COOH) in the molecule C = (CC(= C)c1cc(C(= O)O)cnc1C(C)CC. (a) RNN-based encoders process sequence data, maintaining a memory of previous inputs to effectively capture sequential patterns of (—COOH). (b) Transformer-based models utilize self-attention mechanisms, enabling them to identify and focus on the (—COOH) group's specific interactions within the molecular sequence. (c) GNN-based architectures employ a message passing strategy, extracting the topological information of (—COOH) within the molecule's graph structure. (d) CNN-based models analyze spatial patterns through convolution layers, identifying sub-images that contain the (—COOH) group. This visualization highlights how each encoder type uniquely processes and interprets the molecular structure for MPP.

memory feature that allows them to maintain context and order in sequences. This capability makes RNNs highly effective for tasks involving sequences data. Nowadays some work uses RNN-based model to analyze 1D molecular data, such as SMILES.

Recently, RNN-based models have been used in an increasing amount of works lately to assess 1D molecular data, including SMILES. Lin et al.[4] first transformed SMILES into sample vectors, which are then processed using bidirectional GRU neural networks to predict molecular properties, illustrating an innovative approach in training models for molecular property prediction. Lv et al.[5] introduced Mol2Context-vec to address the challenge of representing molecular substructures and their polysemous nature, integrating different internal state levels for dynamic representations. To highlight the SMILES characters that are more important for the prediction tasks, Wu et al.[50] utilized the bidirectional LSTM attention network in which they employed a novel multi-step attention mechanism to facilitate the extracting of key features from the SMILES strings. Nazarova et al.[61] used the single-layer Elman RNN to identify correlations between the structure of polymers of the norbornene class and their permittivity, while using the SMILES notation in binary and decimal representations. Wang et al.[62] employed a tree-structured LSTM network with signature descriptors to automatically generate expressive signatures for molecular structures, enabling the efficient representation of their structural information and connectivity in a single-step process.

These works demonstrate the effectiveness of RNN in extracting semantic information from SMILES sequences, paralleling methods in Natural Language Processing (NLP). However, they face challenges when incorporating varied expert knowledge and managing long SMILES sequences, and the focus of RNN-based models on adjacent characters hampers effective interactions between distant atoms. This limitation can affect their ability to capture extensive structural relationships, especially when important atoms within the same functional group are distantly placed in the sequence.

4.1.2 GNN-based method

Molecules can be effectively represented as graphs, with atoms as nodes and chemical bonds as edges. GNNs are well-suited to learn from this representation, utilizing layers that enable message passing. In GNNs, node embeddings are updated by aggregating information from neighboring nodes, allowing the network to capture molecular features through atom-level interactions. This method provides a detailed understanding of molecular structures by considering both individual atomic characteristics and their interconnections within the molecule. Yang et al.[51] constructed molecular encodings by using convolutions centered on bonds instead of atoms, thereby avoiding unnecessary loops during the message passing phase of the algorithm. AttentiveFP[10] characterizes the atomic local environment by propagating node information across varying distances, enhancing the representation of molecular structures. Additionally, it incorporates nonlocal intramolecular effects through the use of a graph attention mechanism. Withnall et al.[63] introduced attention and edge memory schemes to the existing message passing neural network framework. To address insufficient bond information extraction, Li et al.[64] explicitly dropped the matrix mapping of edge features and employed a triplet message mechanism. This mechanism calculates messages from atom-bond-atom information and updates the hidden states of neural networks. Zhang et al.[65] proposed CoAtGIN, which uses k-hop convolution to capture long-range neighbor information at the local level and utilizes linear attention to aggregate the global graph representation according to the importance of each node and edge at the global level.

But these methods focus on atom (node) or bond (edge) information. To address this issue, Song et al.[8] proposed a communicative message passing neural network to improve molecular embedding by strengthening the message interactions between nodes and edges through a communicative kernel. SC-NMP[66] aggregates the node representations of the current step and the graph representation of the previous step, and proposes densely self-connected neural message passing, which connects each layer to every other layer in a feed-forward fashion. To extract useful interactions between a target atom and its neighboring atomic groups, Li et al.[67] proposed a new graph learning paradigm based on a block design named block-based GNN, which demonstrates that the network degradation problem can be reduced by applying a block design with normalization and skip-connection. Ma et al.[68] employed cross-dependent

message passing strategy to integrate the node-centered and edge-centered encoders. Liu et al.[69] developed a hypergraph-based topological framework to characterize detailed molecular structures and interactions at the atomic level. They have recently proposed embedding homology and persistent homology. Feng et al.[70] transformed each molecular graph into a heterogeneous atom-bond graph to fully utilize the bond attributes, and designed unidirectional position encoding for such graphs. Biswas et al.[31] passed additional atomic and molecular features, including 2D RDKit descriptors, Abraham parameters, QM descriptors, and 3D geometries, to improve the model performance. Hasebe[71] proposed a knowledge-embedded message passing neural network that can be supervised together with nonquantitative knowledge annotations by human experts on a chemical graph. This graph contains information on the important substructure of a molecule and its effect on the target property. Yang et al.[72] extracted physical information with a neural physical engine that learns molecular conformations by simulating molecular dynamics with parameterized forces. They then employed this physical information as supplementary data for predicting molecular properties.

However, most methods essentially attribute predictions to individual nodes, edges, or node features. This kind of interpretability is only partially compatible with chemists' intuition at best. Chemists are more accustomed to comprehending the causal relationship between molecular structures and properties in terms of chemically meaningful substructures, such as functional groups, rather than individual atoms or bonds. Zang et al.[73] decomposed the molecular graph by BRICS and additional decomposition to construct a motif-level graph, in which corresponding multi-level generative and predictive tasks are designed as self-supervised signals. As the graph pooling technique for learning expressive graph-level representation is critical yet still challenging, Liu et al.[74] proposed master-orthogonal attention, a novel cross-level attention mechanism specifically designed for hierarchical graph pooling. To fully explore higher-order substructure information, Gao et al.[75] proposed substructure interaction attention, which takes both the information of neighbors' substructures and the interaction information among them into account during the aggregation process. To retain locality and linear network complexity, Bouritsas et al.[7] employed a topologically-aware message passing scheme based on substructure encoding, which does not attempt to adhere to the Weisfeiler-Leman hierarchy. Addressing the oversmoothing problem in multi-hop operations, Ye et al.[76] constructed a composite molecular representation with multi-substructural feature extraction and processed such features effectively with a nested convolution plus readout scheme to capture interacting substructural information. Zhu et al.[77] utilized corepresentation learning of molecular graphs and chemically synthesizable BRICS fragments. Furthermore, a plug-and-play feature-wise attention block is first designed in the model architecture to adaptively recalibrate atomic features after the message passing phase. To accurately model the complex quantum interactions inherent in molecules, Lu et al.[78] utilized a sophisticated hierarchical graph neural network, which directly extracts features from both the conformation and spatial information of molecules, and then integrates these features through multilevel interactions. Fey et al.[79] took in two complementary graph representations: the raw molecular graph representation and its associated junction tree, where nodes represent meaningful clusters in the original graph. Focusing on the molecular hierarchical relationship, Han et al.[6] proposed a simple yet effective rescaling module, called contextual self-rescaling, that adaptively recalibrates molecular representations by explicitly modeling interdependencies between atom and motif features. Ji et al.[52] modeled a molecule as a heterogeneous graph and leveraged metapaths to capture latent features for chemical functional groups. They also designed a hierarchical attention strategy to aggregate heterogeneous information at both the node and relation levels. To extract functional groups as motifs for small molecules, Wu et al.[80] constructed a heterogeneous molecular graph with both atom-level and motif-level nodes, and adopted a heterogeneous self-attention layer to distinguish the interactions between multi-level nodes.

Since different 3D structures may lead to dissimilar molecular properties despite having the same 2D molecular topology. Recently, many works utilizing molecular 3D structures have been introduced. To emphasize equivariant constraints, Fuchs et al.[81]

utilized the explicit increase of equivariance constraints in self-attention mechanisms. As rotation-invariant representations struggle to convey directional information, Schutt et al.[82] proposed rotationally equivariant message passing, exemplified by the polarizable atom interaction neural network architecture. Brandstetter et al.[83] expanded equivariant graph networks to include not only invariant scalar attributes but also covariant information, like vectors or tensors. This model consists of steerable MLPs, capable of incorporating geometric and physical information within its message passing and updating functions. Gasteiger et al.[84] showed the universality of spherical representations, employed a two-hop message passing mechanism with directed edge embeddings for rotationally equivariant predictions, and utilized symmetric message passing, augmented with geometric information, to enhance our model's efficacy in MPP. Gasteiger et al.[85] integrated directional information and interatomic distances by embedding and updating messages between atoms, using a spherical 2D Fourier-Bessel basis to jointly represent distances and angles. To model angular relationships among neighboring atoms in a GNN, ensuring constraints like rotation invariance and energy conservation, Shuaibi et al.[86] utilized a per-edge local coordinate frame and innovated a spin convolution, thereby securing rotation invariance in edge messaging. Fang et al.[53] proposed a self-supervised framework using molecular geometric information by constructing a new bond angle graph, where the chemical bonds within a molecule are considered as nodes and the angle formed between two bonds is considered as the edge.

The GNN-based model section concludes by recognizing that while GNN excel in capturing molecular topological information and integrating domain knowledge, their effectiveness is hindered by the small-world phenomenon. This characteristic leads to over-smoothing in deeper networks, where nodes lose feature distinctiveness, impacting predictive accuracy. Additionally, the specialized structure of GNN makes it challenging to scale up with increased parameters, limiting their capability to handle large molecular datasets effectively.

4.1.3 Transformer-based method

Originally excelling in NLP, the Transformer architecture is renowned for its self-attention mechanism, which allows for parallel processing of entire sequences. This capability enables it to efficiently manage long-range dependencies within data, making it highly effective in MPP. Its adeptness at understanding detailed contextual relationships enhances the accuracy and computational efficiency in predictive modeling.

Wang et al.[87] and Chithrananda et al.[55] used Transformer to extract molecular information from SMILES, which is treated as natural language. Wang et al.[88] proposed two significant advances in molecular data processing: structural fingerprint tokenization for more efficient molecule graph tokenization and normalized graph raw shortcut-connection to enhance latent representations in complex model structures. To address challenges in the validity and robustness of SMILES representations, Yüksel et al.[56] uniquely utilized SELFIES, a robust and flexible molecular representation format, to learn high-quality molecular features, enhancing the reliability of molecular data analysis in computational chemistry. To predict activity coefficients in binary mixtures, Winter et al.[89] integrated information from two SMILES strings representing the mixture components, along with temperature and token position data, into a unified matrix for input encoding. Ross et al.[12] delves into the differences between absolute and relative position embeddings in SMILES representation, proposing an efficient linear attention approximation for the RoFormer model[90], which focuses on relative positioning, to enhance molecular SMILES processing in deep learning applications.

The Transformer architecture, originally designed for sequence data, has been effectively adapted for molecular graph representation in recent research. Its proficiency in handling global molecular information enhances its utility in molecular property prediction, showcasing its versatility beyond traditional sequence analysis. Maziarka et al.[91] proposed the molecule attention Transformer, which adapts the Transformer architecture, augmenting the self-attention mechanism with inter-atomic distances and molecular graph structure. Li et al.[9] focused on chemical bonds in molecular representations, employing molecular line graphs to illustrate edge adjacencies in original molecular graphs. Each graph is augmented with a knowledge node containing molecular descriptors and fingerprints, connected to its original nodes. Rong et al.[11] combined message passing networks with a

Transformer-style architecture, extracted vectors as queries, keys, and values from nodes of the graph, then fed them into the attention block. Park et al.[92] introduced graph relative positional encoding, which effectively encodes graph structures by concurrently addressing node-topology and node-edge interactions, bypassing the need for linearization. Hussain et al.[93] developed the edge-augmented graph Transformer, employing global self-attention rather than traditional static convolutional aggregation. This design facilitates dynamic and long-range node interactions, and incorporates edge channels for evolving structural information, enabling direct predictions on edges and links. Masters et al.[94] integrated a substantial message-passing module with a biased self-attention layer to facilitate both localized biases and broad-scale communication. Chen et al.[95] proposed graph propagation attention, which explicitly handles node-to-node, node-to-edge, and edge-to-node interactions, allowing for comprehensive information propagation. Yin and Zhong[13] developed a method that alternates between GNN and Transformer layers, and repeats in sequence. This approach effectively blends local and global information, allowing the graph Transformer to comprehensively integrate node data from both nearby and distant sources. To extrace the coarse-grained view, Ren et al.[96] made the molecular graph first enter the message passing phase of the traditional GNN layers to update the node embeddings, then enter graph transformation layers to learn different granular information. To achieve one encoder for extracting 2D or 3D information, Luo et al.[14] used two separated channels to encode 2D and 3D structural information and incorporated them with the atom features in the network modules. To fully leverages chemical knowledge, Gao et al.[97] constructed an embedding unit comprising a GNN and a Transformer to balance the neighboring and distant interactions of an atom, and more attention is given to conjugated systems, unsaturated bonds, heteroatoms and the molecular topology. To extract the molecular fragment information, Jiang et al.[98] designed a pharmacophoric-constrained multi-view molecular representation graph, enabling PharmHGT to extract vital chemical information from functional substructures and chemical reactions.

Transformer have demonstrated effectiveness in recent work, particularly with sequence data like SMILES, where they treat it similarly to natural language. Their global information extraction capabilities also extend to molecular graph representation. Recent innovations combine Transformer with GNN, enabling simultaneous local and global data analysis. This blend showcases Transformer's strength in handling large molecular datasets and extracting comprehensive insights, which is vital in MPP.

4.1.4 CNN-based method

CNN, known for processing grid-like topology data, is adept at extracting features through convolutional layers and efficiently detecting local patterns. This makes it highly effective for image and pattern recognition tasks, which is a trait utilized extensively in MPP.

To extract the local pattern of 1D molecular data, Hirohara et al.[99]'s innovative application of CNNs to SMILES data for chemical motif detection marks a significant step in computational drug discovery. Chen and Tseng[16] highlighted the impact of SMILES molecular enumeration on CNNs' performance in solubility prediction.

As DL methods have achieved great success in the image processing field, some work use CNN to extract 2D molecular image, but the size of the same atom/structure is vibrational in different molecules, because of the fixed size of the whole molecular image. To address this issue, Zhang et al.[57] introduced ABC-Net, predicting graph structures by representing atoms and bonds as points, utilizing CNN-generated heat-maps. Jiang et al.[100] proposed an equal-sized molecular persistent spectral image, and encodered it with CNN model to extract molecular representation.

As the visual representation of molecular structure, 3D molecular grid is important for extracting molecular 3D information. However, a direct 3D representation of a molecule with atoms localized at voxels is too sparse, which leads to poor performance of the CNNs. To address this issue, Kuzminykh et al.[101] presented a novel approach where atoms are extended to fill other nearby voxels with a transformation based on the wave transform. Liu et al.[17] utilized an atom-centered Gaussian density model for 3D molecular representation, which involves defining multiple channels for different spatial resolutions corresponding to each atom type. Sunseri and Koes[44] facilitated the use of grid-based molecular representations in DL,

generating 3D arrays of voxelized molecular data compatible with various DL frameworks.

The research we have reviewed indicates that the CNN-based networks excel at encoding pixel-based data, like 2D images and 3D grids, understandable to humans. This ability of CNN to efficiently extract local and global information from such data is essential for analyzing molecular behaviors.

4.1.5 Multi-modality-based method

Multi-modal learning, initially prominent in computer vision, is now widely applied in various fields for its ability to handle and integrate different data types. Its key benefit is enhancing model robustness by using complementary data sources. In the field of molecular property prediction, this method has gained popularity.

Due to the significant local chemical information contained in fingerprints may assist models to achieve superior results, Cai et al.[102] and Wang et al.[103] termed fingerprints and graph neural networks, which combine and simultaneously learn information from molecular graphs and fingerprints for MPP. Not only fingerprint, work in Ref. [104], MolFM[105], work in Ref. [106], GraSeq[21], and GIT-Mol[107] employ different encoders to process information from SMILES strings and molecular graphs. Tang et al.[108] encoded molecule by using molecular descriptors and fingerprints, molecular graph, and SMILES text notation. Liu et al.[58] combined molecular structural data and textual knowledge to enhance molecular comprehension, jointly learning the chemical structures of molecules and textual knowledge. Zhang et al.[109] used molecular mass spectrum as another representation to provide supplement information, which is not contained in the graph data. To address neglects 3D stereochemical information, Chen et al.[110] proposed an algebraic graph-assisted bidirectional Transformer framework by fusing SMILES and algebraic graph representations. By broad learning of many molecular descriptors and fingerprint features, MolMap[111] is developed for mapping these molecular descriptors and fingerprint features into robust two-dimensional feature maps. To integrate the 3D coordinates information, Zhou et al.[54] employed the atom distance matrix as the position encoding. Liu et al.[112] incorporated comprehensive relational data, including distance, angle, and torsion information between atoms, extending beyond the traditional edge-based 1-hop interactions. Wang et al.[113] embedded both molecular graphs and sequences, then created a joint embedding space alongside modality-specific spaces to ensure that the multi-modal data maintain both its distinctive characteristics and a consistent representation across different modalities.

In conclusion, the above work underscores the effectiveness of multi-modal learning in the context of MPP. This approach facilitates the seamless integration of various molecular modality, including sequences, graph data types, and molecular descriptors. By amalgamating these diverse sources of information, multi-modal learning provides a richer and more nuanced understanding of molecular properties, which is essential for achieving accurate predictions.

4.2 Training strategy

In this section, we introduce all approaches used to train DL models. While supervised learning has been traditionally predominant, its reliance on scarce labeled data presents limitations. To circumvent this, recent approaches have shifted towards unsupervised, self-supervised, and semi-supervised learning methods, capitalizing on the abundance of unlabeled data. Transfer learning is also employed to utilize models pretrained on unrelated data, enhancing the model's performance on specific tasks. Additionally, multi-task learning strategies are adopted to leverage related labeled data, further refining the model's accuracy in predicting molecular properties. As Figs. 6 and 7 shown, the details of training strategy are as followings.

4.2.1 Self-supervised learning

Self-supervised learning, widely used in NLP[124, 125], utilizes unlabeled data to extract prior knowledge, proving effective in addressing labeled data scarcity. This method empowers models to learn comprehensive representations from abundant unlabeled data, enhancing their learning capabilities and insight extraction.

Inspired by NLP, Wang et al.[87], Chithrananda et al.[55], Zhang et al.[126], Ahmad et al.[127], and Irwin et al.[128] employed Masked Language Modeling (MLM) on large scale unlabeled data to generate context-sensitive representation, treating SMILES as natural language. Ma et al.[3] used auto-encoder strategy in pretrain stage, first converted SMILES to a vector representation, and then reconstructed representation back to SMILES to update the network. Furthermore, Guo et al.[21] fused the molecular graph

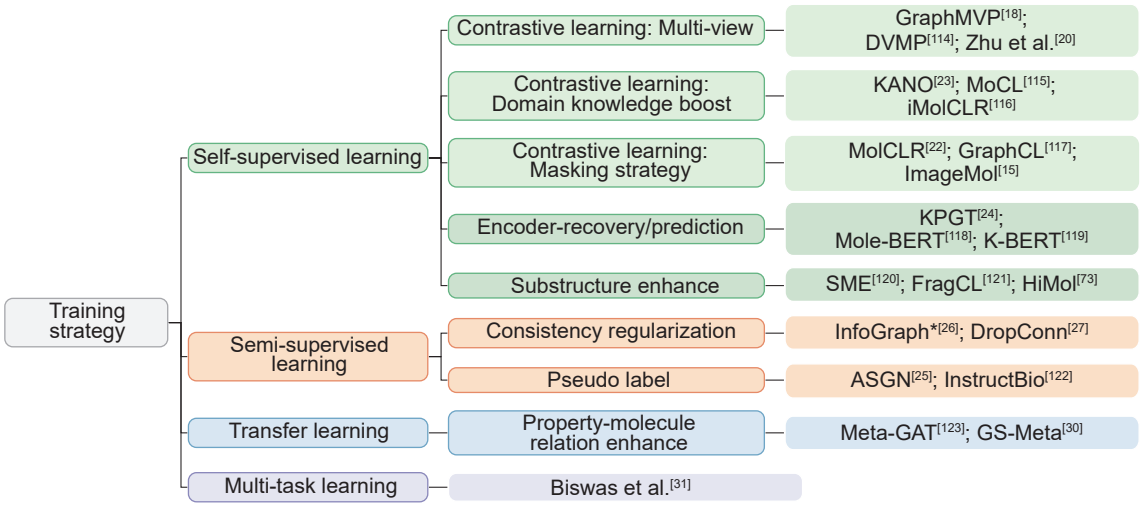

Fig. 6 Training strategy summary. We categorize training strategies into four key types: self-supervised learning, semi-supervised learning, transfer learning, and multi-task learning. Each category includes a detailed description of the main focuses and considerations prevalent in renowned studies, illustrating the diverse approaches and priorities within each training strategy for optimizing molecular property prediction.

and SMILES representation to reconstruct the SMILES. Except using SMILES as input, Yuksel et al.[56] employed MLM in SELFIES representations in order to obtain their concise, flexible, and meaningful representations. To let nodes appear in similar structural contexts to nearby embeddings, Hu et al.[129] proposed a context prediction task by using subgraphs to predict their surrounding graph structures. To address GNN oversmoothing and encourage latent node diversity, Godwin et al.[130] employed denoise technique, in which they corrupted the input graph with noise, and added a noise correcting node-level loss. Zeng et al.[15] implemented an auto-encoder for molecular image reconstruction, using a discriminator to distinguish between real and fake molecular images. To expand atom vocabulary, Xia et al.[118] used a context-aware tokenizer to encode atom attributes into meaningful discrete codes, then randomly masked and recovered these codes to efficiently pretrain their encoder. Intrinsically, for molecules, a more natural representation is based on their 3D geometric structures, which largely determine the corresponding physical and chemical properties. To overcome the challenge of attaining the coordinate denoising objective, Liu et al.[131] employed an SE(3)-invariant score matching strategy to successfully transform such objective into the denoising of pairwise atomic distances. To capture the anisotropic characteristics of molecules, Feng et al.[132] proposed a novel hybrid noise strategy, including noises on both dihedral angel and coordinate, and also decoupled the two types of noise and designed a novel fractional denoising method, which only denoise the latter coordinate part. For effectively learning 3D spatial representation, Zhou et al.[54] employed 3D position recovery and masked atom prediction as pretrain task. Furthermore, Jiao et al.[133] exploited the Riemann-Gaussian distribution to ensure the loss to be E(3)-invariant, enabling more robustness. To guild by the molecular domain knowledge and extract chemical information like chemists, Li et al.[9, 24] leveraged the molecular descriptors and fingerprints, which serves as the semantics lost in the masked graph to guide the prediction of the masked nodes, thus making the model capture the abundant structural and semantic information from large-scale unlabeled molecules. Wu et al.[119] proposed atom property prediction to discern finer differences between atoms, and MACCS fingerprints prediction, enabling their model to extract and learn predefined molecular features. Gao et al.[134] used atom charges and 3D geometries as inputs, with molecular energies as the target labels, aiming to effectively leverage energy information for enhanced molecular analysis. To optimize multi-task integration and avoid ineffective transfer, Wang et al.[135] introduced a fusion strategy that utilizes a surrogate metric based on the total energy of all atoms in a molecule during the pretraining stage. Zang et al.[73] designed three generative tasks that predict bond links, atom types, and bond types with the atom

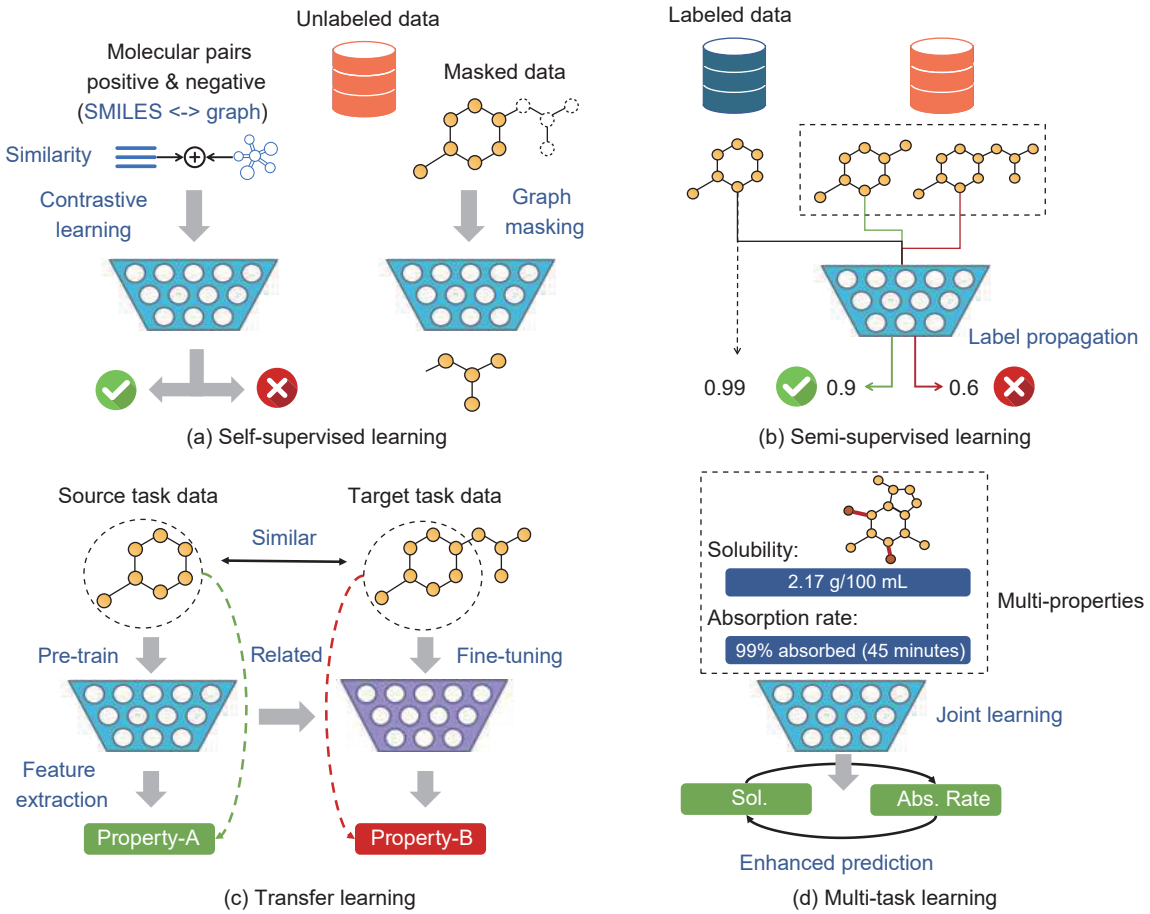

Fig. 7 Training strategy. The high cost of experiments often results in a scarcity of labeled data, leading to challenges like overfitting and poor generalization. To address this, this figure illustrates various advanced training strategies. (a) Self-supervised learning, such as contrastive learning and masking recovery, utilizes unlabeled data for pretraining encoders, enabling them to learn molecular domain knowledge effectively. (b) Semi-supervised learning methods, like pseudo-labeling and co-training, leverage both labeled and unlabeled data, enhancing the encoder's understanding of data distribution. (c) Transfer learning strategy capitalizes on pre-trained models from source tasks to boost performance on target tasks. (d) Multi-task learning approaches combine related datasets to predict multiple properties simultaneously, benefiting from the relational aspects of molecular data.

representations, and designed two predictive tasks that predict the number of atoms and bonds with the molecule representation. Zeng et al.[136] and Broberg et al.[137] developed methods to predict the product molecular SMILES based on the reactant molecular embedding. This approach allows for the extraction of chemical information from chemical reactions, providing insights into the molecular transformations involved in the reaction process.

Contrastive learning, a method distinguishing positive and negative molecule pairs, has become a key strategy in encoder pretraining for its ability to enhance molecular structure discernment. This technique is extensively utilized in numerous studies, making it a cornerstone for improving molecular structure recognition in various models. For the SMILES augmentation, Wu et al.[50] and Zhang et al.[138] implemented SMILES enumeration, a technique that varies starting atoms and traversal orders to represent a molecule with different SMILES, thereby uncovering more intricate patterns from complex SMILES structures. Wu et al.[119] and Abdel-Aty and Gould[139] utilized SMILES permutation as a data augmentation technique, involving the rearrangement of atoms in an SMILES string to create different representations without altering the underlying molecular structure. For molecular graph augmentation, techniques like node dropping, edge perturbation, attribute masking, and subgraph masking are commonly used[22, 117, 140, 141]. However, these random masking methods may not

effectively guide the encoder to identify the most crucial chemical information, and might result in the creation of less accurate positive and negative molecule pairs for the training process. To capture important molecular structure and higher order semantic information, Liu et al.[142] adopted the graph attention network as the molecular graph encoder, and leveraged the learned attention weights as masking guidance to generate molecular augmentation graphs. Lin et al.[143] first modeled the underlying semantic structure of the graph data via clustering semantically similar graphs to select the positive and negative pair, and then reweighted its negative samples based on the distance between their prototypes and the query prototype, such that those negatives having moderate prototype distance enjoy relatively large weights. Cui et al.[144] utilized the GNN encoder and its momentum-update version[145] to generate positive samples at the representation level, and selected the negative pairs by the semantic importance of nodes, which is calculated by eigenvector centrality iteration[146]. Wang et al.[147] employed a generative probabilistic model to learn molecular graph structures for topology augmentations, and simultaneously developed feature selectors to mask less critical atom features, thus generating effective attribute-level augmentations. To gain deeper insights into chemical information, many researchers incorporate domain knowledge into their contrastive learning approaches. By using backbone and side-chain information, Liu et al.[148] employed side-chain repetition, side-chain generation, backbone disruption, and backbone disruption+side-chain deletion strategy to generate hard positive, soft positive, soft negative and hard negative samples, respectively.

Sun et al.[115] replaced a valid substructure by a bioisostere that introduces variation without altering the molecular properties too much, and treats them as positive pairs. Also, they optimized the similarity of molecule pairs embedding to be close to the similarity of their ECFP. To avoid faulty negative pairs, Wang et al.[116] mitigated negative contrastive instances by considering ECFP similarities between molecule pairs. Wang et al.[149] calculated the weight vector using the self-attention mechanism to determine the selection probability of each character in SMILES, and generated positive samples using three masked strategies: roulette masking, top masking, and random masking. To maintain semantics between conformers, Moon et al.[150] randomly selected molecules from the conformer pool instead of selecting the most stable molecules to learn the 3D structure abundantly. Kuang et al.[151] considered conformations with the same SMILES as positive pairs and the opposites as negative pairs, while keeping the weight to indicate the 3D conformation descriptor and fingerprint similarity. Knowledge Graph (KG) is a semantic network composed of entities and their relations in the real world[152]. Hua et al.[153] used the atoms in SMILES as indices to query the embedding matrix to obtain entity and relation embeddings. For the entity and relation vectors of different atoms, they obtained the entity and relation embeddings of the SMILES through linear mapping, and finally concatenated the two vectors to obtain the final embedding representation. Fang et al.[154] first constructed a chemical element KG based on periodic table of elements, to describe the relations between elements and their basic chemical attributes, Furthermore, Fand et al.[23] constructed another chemical element KG based on the periodic table and Wikipedia pages to summarize the basic knowledge of elements and the closely related substructure. Those KGs offer a comprehensive and standardized view from a chemical element perspective, and help to augment the original molecular graph with the guidance of KG.

In the realm of MPP, a deep understanding of molecular substructures is increasingly recognized as crucial. Many recent studies leverage this domain knowledge to effectively identify and analyze important substructural information, significantly enhancing the understanding of molecular behavior. Xu et al.[155] aimed to preserve local similarities between graph instances by aligning embeddings of related subgraphs and differentiating these from unrelated pairs. They also implemented hierarchical prototypes to represent the latent distribution of graph datasets, enhancing data likelihood with respect to both GNN parameters and these hierarchical structures. Xu et al.[156] aimed to capture both intra- and inter-space relationships in group representations, and also introduced an attention-based mechanism for generating groups, aiming at identifying diverse molecular substructures effectively. Wang et al.[116] employed BRICS to decompose different substructures, which are considered as contrastive negative pairs. Motifs, including chemical functional

groups or fragments, serve as self-generated labels determined by their presence or absence in the graph. Shen et al.[157] and Rong et al.[11] used these labels to pretrain their encoders. To learn the local semantics, Luo et al.[158] used graph clustering techniques to partition each whole graph into several subgraphs, while preserving as much semantic information as possible, and treated the molecular graph and the clustering graph as positive pair. Benjamin et al.[159] extracted substructure information by setting the junction tree (through a tree decomposition algorithm) reconstruction and fingerprint prediction task. To analyze molecular GNN strictly in terms of chemically meaningful fragments, Wu et al.[120] identified the most crucial set of substructures (BRICS and Murckoand functional groups) in a molecule that are responsible for a model's prediction. HeGCL[160] introduces the meta-path view that provides semantic information, and encodes graph embeddings by maximizing mutual information between global and semantic representations obtained from the outline and meta-path view, separately. Hierarchical molecular graph is a usual way to extract the substructure molecular representations. Zhu et al.[77] extracted hierarchical information by utilizing co-representation learning of molecular graphs and chemically synthesizable BRICS fragments, and also used a feature-wise attention block to adaptively recalibrate atomic features after the message passing phase. Kim et al.[121] constructed a bag of fragments from a molecule through fragmentation, treating a complete or incomplete bag as a positive or negative view of the original molecule, respectively. Xie et al.[161] proposed a Fragment-based Molecular Graph (FMG) to represent the topological relationship between chemistry-aware substructures within a molecule. They then pretrained it on a fragment level using contrastive learning with well-designed hard negative pairs to extract node representations in FMGs. Ji et al.[162] decomposed the molecular graph using a more reasonable method to construct the fragment graph. They selected positive/negative pairs based on similarities between two-level molecule pairs, and employed a contrastive loss function, as proposed by Hadsell et al.[163], to pretrain the encoder.

Diverse data formats have been shown to be crucial for MPP, and the multi-modal approach, merging these formats, enhances prediction accuracy by offering a holistic view of molecules. This technique, increasingly adopted in research, combines different data types for a more detailed molecular analysis. To leverage two popular molecular representations and augmentations for each modality, Pinheiro et al.[164], Zhang et al.[165], Zhu et al.[114], and Sun et al.[106] exploited two molecular representations that can be easily acquired from chemical space: the SMILES string and the molecular graph, and then made them as positive pairs. Li et al.[19] utilized self-supervised learning by exploiting the relationship and consistency between 2D topological and 3D geometric structures of molecules. Additionally, Liu et al.[18] applied a generative self-supervised learning approach that focuses on intra-data knowledge, reconstructing key features at the individual data point level to enhance the understanding of molecular structures. 3D Infomax[166] maximizes the mutual information between learned 3D summary vectors and the representations of a GNN. Zhu et al.[20] implemented a multifaceted pretraining strategy involving the reconstruction of masked atoms and coordinates, generating 3D conformations based on 2D graphs, and creating 2D graphs from 3D conformations. Kim et al.[121] focused on extracting explicit 3D geometric information by proposing a solution for predicting torsional angles between adjacent molecular fragments, thereby enhancing the depth and accuracy of 3D molecular analysis. Zhu et al.[167] aimed to integrate multiple molecular feature views, including 2D and 3D graphs, Morgan fingerprints, and SMILES strings, ensuring cohesive embedding consistency between these among representations for a more unified molecular analysis.

The reviewed works show that self-supervised learning, particularly through methods like encoder-recovery and contrastive learning, effectively utilizes unlabeled data to improve model generalization in MPP. These methods excel in learning prior knowledge through various pretraining tasks, allowing for integration of multi-modal data and domain knowledge. This approach significantly enhances the adaptability and performance of models in molecular property prediction scenarios.

4.2.2 Semi-supervised learning

Semi-supervised learning effectively alleviates the scarcity of labeled molecular data in fields like MPP. By blending a small subset of labeled data with a larger

pool of unlabeled data, it bridges the gap between fully supervised and unsupervised learning methods.

Consistency regularization is based on the idea that applying realistic perturbations to unlabeled data should not significantly alter predictions, ensuring stability and reliability in the learning process. InfoGraph*[26] employs mean-teacher method[168] to maximize the mutual information between unsupervised graph representations and the representations learned by existing supervised methods in semi-supervised scenarios. Chen et al.[169] predicted chemical toxicity and trained the network by the mean-teacher SSL algorithm, which updates the weights in teacher model by applying the exponential moving average. Zhang et al.[27] proposed a data augmentation, which constructs new adjacency matrix and randomly masks the edges, calculates the average of all data augmentation distributions, and then employs MixMatch[170] label guessing and sharpening method to minimize entropy and accurately guess labels based on the label distribution center.

Proxy-label strategy, assigning temporary labels to unlabeled data, expands the training dataset when labeled data are limited. This approach enhances the model's learning process, with the proxy labels being iteratively refined for improved accuracy and generalization. ASGN[25] adopts a teacher-student framework to jointly exploit information from molecular structure and molecular distribution to learn general representation, then employs the active learning strategy in terms of molecular diversities to select informative data. Yu et al.[171] developed a semi-supervised drug embedding model, that combines unsupervised learning from the chemical structures of drugs and drug-like molecules with supervised learning based on hierarchical relations from an expert-crafted drug hierarchy. This approach ensures a robust and comprehensive representation of drug properties. Ma et al.[172] employed teacher-student framework, which used several epochs as an iteration, updating teacher model by the best student model. As the Cross-Entropy (CE) loss function is not proved to be robust to label noise during the training, they employed generalized CE[173] loss to boost the self-training. To address data imbalance, Liu et al.[174] analyzed the distribution of imbalanced annotated data and identified label ranges needing adjustment, and then used high-quality pseudo-labels to create graph examples to augment under-represented areas, striving for an ideal balance in training data. Wu et al.[122] introduced an instructor model to provide the confidence ratios as the measurement of pseudo-labels' reliability. These confidence scores then guide the target model to pay distinct attention to different data points, avoiding the over-reliance on labeled data and the negative influence of incorrect pseudo-annotations.

This approach not only enhances model performance by utilizing the comprehensive information available in unlabeled data, but also addresses the challenge of acquiring extensive labeled datasets, which is a common issue in MPP.

4.2.3 Transfer learning

Transfer learning strategies, widely adopted in various fields to address data scarcity, focus on enhancing prediction performance for tasks with limited data[175–177]. These strategies involve transferring knowledge from a data-rich source task to improve learning in a data-scarce target task. Recently, there has been a significant increase in methods employing transfer learning, showcasing its growing importance and application across different domains.

Sun et al.[28] enhanced chemical and physiological property predictions by applying transfer learning, integrating insights from physics and physical chemistry to improve training outcomes. Li et al.[178] developed a framework for accurately estimating task similarity, which, as demonstrated in comprehensive tests, provides valuable guidance for enhancing the prediction performance of transfer learning in molecular property analysis.

Meta-learning, focusing on rapid adaptation to new tasks with minimal data, is effective in addressing the lack of labeled molecular data. Many recent works[179–183] are based on Model-Agnostic Meta-Learning (MAML), enabling rapid adaptation and learning in data-limited scenarios. To effectively utilize correlations of molecules and properties, Lv et al.[123] constructed a molecule-property relation graph, where nodes represent molecules and properties connected by property labels, and then redefined a meta-learning episode as a subgraph within it, containing a target property node along with related molecules and auxiliary property nodes. Chen et al.[29] developed ADKF-IFT, a model that separately trains a subset of parameters with meta-learning loss and adapts others using maximum marginal likelihood for each task. This

method, unlike previous ones using a single loss for all parameters, effectively utilizes meta-learning's regularization to prevent overfitting. MTA[184] is mainly conducting task augmentations by generating new labeled samples through retrieving highly relevant motifs from a pre-defined motif vocabulary as an external memory. To utilize many-to-many correlations of molecules and properties, Zhuang et al.[30] constructed a Molecule-Property relation Graph (MPG), reformulated an episode in meta-learning as a subgraph of the MPG, and then scheduled the subgraph sampling process with a contrastive loss function, which considers the consistency and discrimination of subgraphs. Guo et al.[185] developed a model where the importance of different property prediction tasks in few-shot learning is gauged using a self-attentive task weight, calculated by averaging molecular embeddings from each task's query set, to represent task significance. Wang et al.[186] proposed a property-aware embedding function for context-based molecular adaptation and an adaptive relation graph module for molecular relation and embedding refinement, and then employed selective meta-learning strategy for task-specific parameter updates, effectively harmonizing shared knowledge and unique aspects in property prediction tasks. Yao et al.[187] picked out some molecules sharing common properties and used multiple property-aware graph neural networks to extract molecular representation, then employed the Spearman's correlation to build property-aware matrix. In the few-shot MPP task, the meta-learning strategy is adopted to learn common prediction knowledge from the meta-training categories.

In conclusion, transfer learning has gained popularity for its ability to enhance model generalization in scenarios with limited labeled data. This method is particularly effective in exploiting the relationships between molecules and properties, identifying shared information, such as the role of molecular substructures across different tasks, which is crucial for developing more informed and accurate predictive models.

4.2.4 Multi-task learning

Multi-task learning is a machine learning approach where a model is trained on multiple related tasks simultaneously, rather than training on each task independently. This strategy leverages the commonalities and differences across tasks, allowing the model to learn more generalizable features.

Ma et al.[3] established a multi-label supervised model on a combined dataset with missing labels. the input to prediction network is a data matrix with multiple property label information, which can be an original dataset collected from specialized experiments. Tan et al.[32] constructed our multitask models by stacking a base regressor and classifier, enabling multitarget predictions through an additional training stage on the expanded molecular feature space. Biswas et al.[31] employed a multitask training method for a single model to predict critical properties and acentric factors, while also adjusting target weights in the loss function to correct data imbalance.

These works we have reviewed show that multi-task learning is highly effective in MPP, as it capitalizes on the interrelation of various molecular properties. This enhances a model's capacity to simultaneously predict multiple properties, which is a particularly valuable trait when dealing with the challenge of limited labeled data.

5 Evaluation and Benchmark

In evaluating the performance of models in MPP, it is crucial to consider a variety of benchmarks, each offering distinct datasets and posing unique challenges. Key benchmarks include MoleculeNet[33], ADMETlab[188], MoleculeACE[189], DrugOOD[190], MD17[191], TUDataset[192] (comprising MUTAG, PTC, NCI, PROTEINS, D&D, and ENZYMES), and PCQM4Mv2[193]; their details are shown in Table 1. MoleculeNet, our primary focus, offers a diverse collection of datasets in quantum mechanics, physical chemistry, biophysics, and physiology, which is crucial for multifaceted molecular property predictions. ADMETlab, is vital for assessing drug safety and efficacy, providing data on ADMET properties. MoleculeACE focuses on QSAR modeling challenges, notably activity cliffs. DrugOOD, based on ChEMBL, emphasizes out-of-distribution generalization in AI-aided drug discovery. MD17 is essential for validating models in computational chemistry with its molecular dynamics trajectories. TUDataset includes varied datasets like DD, ENZYMES, PROTEINS, and MUTAG, each presenting unique graph-based bioinformatics challenges. Lastly, PCQM4Mv2 from the Open Graph Benchmark offers large-scale quantum mechanical property prediction challenges for graph neural network models. Among these, MoleculeNet

Table 1 Overview of datasets for molecular property prediction. This table encapsulates key benchmarks, highlighting their scale, scope, and specific applications in the fields of molecular modeling, drug discovery, and computational chemistry.

Benchmark	Description	Number of molecules	Application/Challenge
MoleculeNet[33]	A diverse collection of datasets across quantum mechanics, physical chemistry, and biophysical properties, pivotal for various molecular property predictions.	785 951	Multifaceted challenges in molecular property predictions
ADMETlab[188]	It provides extensive data on ADMET properties, which is crucial for drug safety and efficacy assessments, enhancing drug development processes.	94 387	Drug development and safety evaluation
MoleculeACE[189]	It focuses on QSAR modeling challenges, especially activity cliffs where minor structural changes cause significant bioactivity variations, testing the robustness of ML models.	48 707	Model accuracy in subtle molecular variations
DrugOOD[190]	Based on ChEMBL, it emphasizes Out-Of-Distribution (OOD) generalization, crucial for advancing AI in drug discovery under limited and varied data scenarios.	930 314	OOD generalization in AI-aided drug discovery
MD17[191]	It contains molecular dynamics trajectories, essential for developing and validating models in computational chemistry and molecular simulations.	3 817 604	Molecular dynamics model development and validation
TUDataset[192]	It includes datasets like DD, ENZYMES, PROTEINS, and MUTAG, each offering unique bioinformatics challenges in graph-based analysis, such as protein structure and enzyme function classification.	–	Bioinformatics applications in graph-based learning
PCQM4Mv2[193]	A dataset from the Open Graph Benchmark, providing large-scale quantum mechanical property prediction challenges for graph neural network models.	3 746 619	Quantum mechanical property prediction in molecular systems

stands out due to its comprehensive coverage and wide usage, making it an exemplary benchmark for our evaluation.

MoleculeNet, a frequently used benchmark in MPP, offers a diverse range of datasets categorized into four groups: quantum mechanics, physical chemistry, biophysics, and physiology. Each group provides specialized datasets to assess different aspects of molecular properties. Quantum mechanics: Datasets in this group are centered around electronic properties derived from quantum mechanical calculations. Physical chemistry: Datasets in this group focus on physical and chemical properties of molecules, including solubility and lipophilicity. Biophysics: This category includes datasets related to biological interactions and processes, such as protein-ligand binding affinities. Physiology: Datasets here pertain to organism-level effects, like toxicity and drug efficacy. Evaluating models across these diverse datasets from MoleculeNet allows for a comprehensive assessment of their predictive capabilities in various aspects of MPP.

Consistent with prior studies, we adopt the area under the receiver operating characteristic curve (namely ROC-AUC) as the evaluation metric for classification datasets, which is a widely used metric for assessing the performance of binary classification tasks. For the regression datasets, we utilize Root-Mean-Squared Error (RMSE) as the evaluation metric. It is important to note that many studies in this field adopt either random or scaffold splits for dividing their datasets, though not uniformly. A random split involves randomly dividing the dataset into training, validation, and test sets, regardless of molecular structures. On the other hand, a scaffold split organizes molecules based on their core chemical scaffolds, ensuring that the model is tested on chemically distinct molecules from those it is trained on, providing a more stringent test of its generalization ability. The choice between these splitting methods can significantly affect the outcomes and interpretations of model performance evaluations.

6 Discussion

6.1 Domain knowledge integration

This part aims to analyze the contribution of domain knowledge for MPP, as the model input. It is divided into 3 part: atom-bond property, molecular structure, and molecular property relation.

As more research utilizes atom and bond properties, the efficiency of MPP has improved. However, it raises the question: Does integrating additional atom and

bond properties into the model input necessarily lead to higher model performance? Wojtuch et al.[194] analyzed the impact of atomic features in graph convolutional neural networks, comparing twelve hand-crafted and four literature-based feature combinations. Findings indicate that feature importance is task-specific and linked to their prevalence in the dataset. Reducing less frequent or redundant features, such as formal charges or aromaticity, improves performance. These insights also apply to advanced models like graph Transformers, though optimal feature selection varies by model.

Increasingly, molecular structure information is being incorporated into MPP, with several studies leveraging it to derive coarse-grained molecular insights from hierarchical graphs. Recent methods, like MoLGNN[157], HiGNN[77], MISU[159], CAFE[161], and iMolCLR[116], have used molecular substructure knowledge, such as BRICS or functional groups, to construct hierarchical graphs treating fragments as nodes. These methods have shown improved results over those not using substructure information. As Tables 2 and 3 demonstrate, the ablation studies reveal a notable enhancement in methodology efficacy when fragment or functional group information is integrated. Specifically, we observe a 3.98% improvement in regression tasks, measured using RMSE, and a 1.72% improvement in classification tasks, measured using ROC-AUC. These results confirm the significant impact of incorporating substructure domain knowledge into these deep learning models. We present two compelling case studies that illustrate the impact of molecular substructure information obtained via the BRICS methodology. The first example involves the molecule "CC(C)(C)NCC(O)c1ccccc1F". When employing BRICS fragmentation, the model identifies the fluorine atom and the tertiary amine as critical features. These fragments are known to

Table 2 ROC-AUC comparison of DL methods for MPP classification tasks with substructure domain knowledge in MoleculeNet. This table contrasts various models, focusing on classification tasks. Each model is evaluated with and without substructure information, as indicated by original and ablation study rows. The "–" symbol marks the absence of data for some datasets, while "avg. imp." shows the average performance improvement due to substructure information integration.

(%)

Model	Splitting	Classification							avg. imp.
		BBBP	Tox21	ToxCast	SIDER	ClinTox	BACE	HIV	
MoLGNN[157] (MoLGNN, only GINVAE)	Random	88.9	–	–	63.6	94.2	87.4	78.0	1.09
		89.2	–	–	61.7	93.7	87.1	76.3	
HiGNN[77] (HiGNN, without HI)	Random	93.2	85.6	–	65.1	93.0	89.0	–	0.25
		93.0	85.2	–	65.4	92.6	88.7	–	
MISU[159] (MISU, without JTVAE)	Scaffold	66.7	76.3	62.8	59.7	78.0	70.5	–	1.93
		65.9	76.2	62.3	58.4	76.1	67.1	–	
CAFE[161] (CAFE-MPP, only Graphormer)	Random	96.5	80.5	–	65.8	98.2	93.9	–	3.93
		93.6	79.3	–	61.8	94.3	89.1	–	
iMolCLR[116] (iMolCLR and MolCLR)	Scaffold	76.4	79.9	73.6	69.9	95.4	88.5	80.8	1.39
		73.6	79.8	72.7	68.0	93.2	89.0	80.6	

Table 3 RMSE comparison of DL methods for MPP regression tasks with substructure domain knowledge in MoleculeNet. This table contrasts various models, focusing on regression tasks. Each model is evaluated with and without substructure information, as indicated by original and ablation study rows. The "–" symbol marks the absence of data for some datasets, while "avg. imp." shows the average performance improvement due to substructure information integration.

Model	Splitting	Regression					avg. imp.
		ESOL	FreeSolv	Lipo	QM7	QM8	
HiGNN[77](HiGNN, without HI)	Random	0.532	0.915	0.549	–	–	2.78%
		0.536	0.941	0.575	–	–	
CAFE[161](CAFE-MPP, only Graphormer)	Random	0.687	1.276	0.684	43.75	0.0141	1.37%
		0.782	1.303	0.718	40.69	0.0138	
iMolCLR[116](iMolCLR and MolCLR)	Scaffold	1.130	2.090	0.640	66.30	0.0170	7.78%
		1.110	2.200	0.650	87.2	0.0174	

significantly affect CNS activity due to their lipophilicity, which is a crucial determinant for Blood-Brain Barrier (BBB) penetration. The second case focuses on CC(C)(O)C(C)(O)c1ccc(Cl)cc1, where BRICS fragmentation reveals the delicate balance between hydrophilic hydroxyl groups and lipophilic chlorinated benzene components. This balance plays a pivotal role in the molecule's ability to penetrate the BBB. Both examples, depicted in Fig. 8, showcase enhanced performance of HiGNN[77] when integrating substructure information, confirming the model's superior ability to predict BBB penetration by capturing intricate substructure information.

Identifying a fundamental set of properties for molecular prediction tasks is crucial for future research. Many studies, including multi-task learning methods, have shown that fundamental molecular properties can enhance other prediction tasks. For instance, Sun et al.[28] improved the training of chemical and physiological property predictors by incorporating related physics property prediction tasks. Additionally, Biswas et al.[31] demonstrated the significance of critical properties and acentric factors, along with four phase change properties as auxiliary targets.

However, the integration of domain knowledge into molecular property prediction models is not without challenges. Firstly, there is still a lot of domain knowledge that is not digitized or gathered, even with the advances in tools like RDKit. It is possible to overlook important subtleties when converting intricate expert information into an electronic format that is easy to use. Secondly, this integration process can introduce biases due to subjective interpretations by domain experts, potentially skewing model outcomes and impacting scalability and adaptability to new molecular data types. Lastly, the requirement to customize deep learning architectures to incorporate such knowledge significantly increases complexity and computational costs, complicating model development and training.

In conclusion, while domain knowledge integration is beneficial, it necessitates a careful and balanced approach. It is crucial to maintain the flexibility, scalability, and objectivity of models. These challenges highlight the need for ongoing efforts to capture and digitize comprehensive domain knowledge, maintaining a critical balance between accuracy and the practical application of these predictive models.

6.2 Multi-modal data utilization

This part aims to analyze the contribution of different modalities in multi-modal models for MPP. It delves into understanding how individual and combined modalities affect prediction performance. Data from various studies are collated and analyzed, emphasizing the contribution of different modalities to MPP tasks, with models in Table 4. Apart from ClinTox, there is uniformity in the predictive prowess displayed by all models across a spectrum of tasks. Nonetheless, ClinTox predictions are prone to biases in

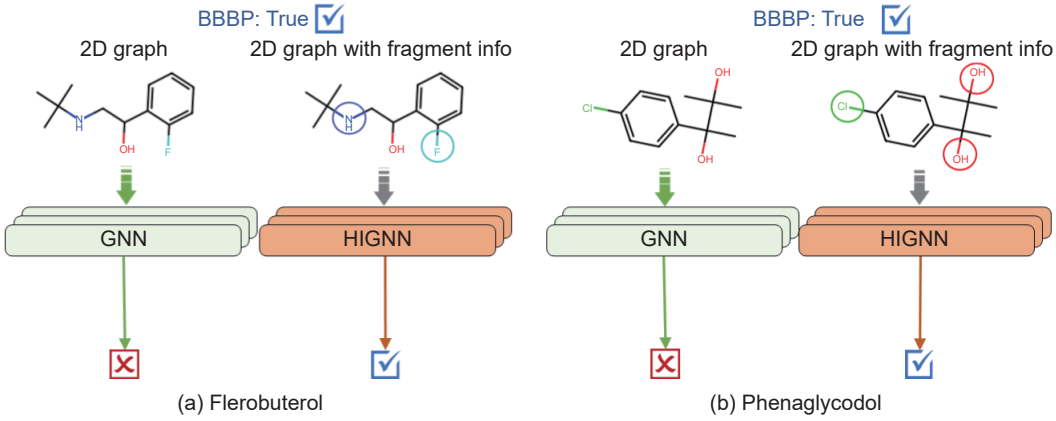

Fig. 8 Case study of molecular fragment information enhance for the BBBP task. This figure compares the predictions of HiGNN[77] for two molecules, Flerobuterol and Phenaglycodol, with and without the use of BRICS-derived molecular substructure information. In (a), Flerobuterol's molecular structure without substructure information leads to an incorrect BBBP prediction, while in (b), incorporating key fragments like the fluorine atom and tertiary amine yields an accurate prediction, highlighting these features' role in CNS activity and BBB penetration. Similarly, for Phenaglycodol, BRICS fragmentation reveals critical hydrophilic and lipophilic components, resulting in a correct BBBP prediction, demonstrating the model's improved predictive capability when domain knowledge is applied.

Table 4 Comparison (ROC-AUC) of multi-modal learning methods in MoleculeNet. The "Method" column specifies the learning strategy: "T" denotes the use of textual data, "S" denotes the use of SMILES data, "2d" and "3d" refer to the use of 2D and 3D molecular graphs, "CL" indicates contrastive learning, and "CA" stands for cross-attention fusion. In the column of "Type of input data", "✓" highlights the types of input data utilized by the models, with the possibility of multiple selections. The best results are emboldened, and the second-best results are highlighted in red.

(%)

Model	Method	Index	Type of input data			Classification						Average
			S	2d	3d	BBBP	Tox21	ToxCast	SIDER	ClinTox	BACE	
KV-PLM [195]	–	0	✓	–	–	72.0	70.0	55.0	59.8	89.2	78.5	70.8
GIN [196]	–	1	–	✓	–	65.4	74.9	61.6	58.0	58.8	72.6	65.2
GraphMVP [18] (GIN, SchNet)	CL (2d, 3d)	2	–	✓	–	68.5	74.5	62.7	62.3	79.0	76.8	71.7
	CL (2d, 3d), CL (2d)	3	–	✓	–	72.4	74.4	63.1	63.9	77.5	81.2	72.1
MoMu-S [197]	CL (S, 2d)	4	–	✓	–	70.5	75.6	63.4	60.5	79.9	76.7	71.1
MoMu-K [197]		5	–	✓	–	70.1	75.6	63.0	60.4	77.4	77.1	70.6
MoleculeSTM [58] (MegaMolBART, GIN)		6	✓	–	–	70.8	75.7	65.2	63.7	86.6	82.0	74.0
		7	–	✓	–	70.0	76.9	65.1	61.0	**92.5**	80.8	74.4
MolFM [105] (KV-PLM, GIN)	CL (S, 2d), CA	8	–	✓	–	72.2	76.6	64.2	63.2	78.6	82.6	72.9
		9	✓	✓	–	72.9	77.2	64.4	64.2	79.7	83.9	73.7
GIT-Mol [107] (SciBERT, MoMu-S)		10	✓	–	–	71.9	73.9	62.1	60.1	83.5	68.4	70.0
		11	–	✓	–	71.1	75.4	65.3	58.2	78.9	65.8	69.1
		12	✓	✓	–	73.9	75.9	66.8	63.4	88.3	81.1	74.9
MolLM [198]	CL (T, 2d), CL (T, 3d)	13	–	✓	✓	**75.7**	**80.0**	**68.2**	**71.0**	91.1	**84.1**	**78.4**

Transformer-based models due to data distribution peculiarities, which results in polarized predictions. Graph-based models like GraphMVP, MoMu, and GIT-Mol(2d) demonstrate a reduction in such bias, albeit with compromised performance in ClinTox.

From an input modality perspective, taking the BBBP task of 2d graph and SMILES information fusion as an example from GIT-Mol[107], the size of the test dataset by scaffold split is 204. We select study cases in which the results are superior to the baseline (SciBERT) after modality fusion, as shown in Fig. 9. This illustration reveals the beneficial role of SMILES modality data in augmenting graph2d data, whereby the integrated representation vectors can rectify erroneous predictions to a certain extent. Conversely, accurate predictions from graph2d, when paired with incorrect SMILES predictions, can also prevent potential mistakes, showcasing the complementary strengths of integrating diverse modalities in enhancing predictive accuracy.

In the examining pre-training strategies, contrastive learning clearly demonstrates significant benefits.

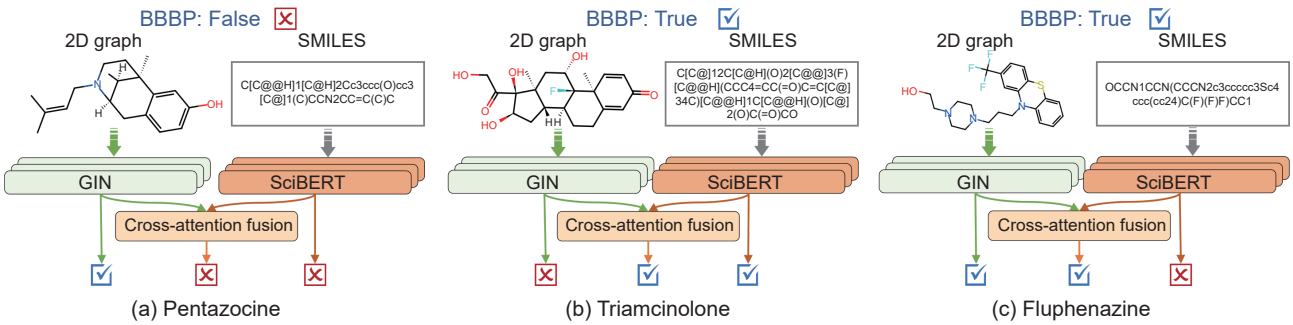

Fig. 9 Case studies of multi-modal fusion for the BBBP Task. Some molecules, such as pentazocine and triamcinolone, may give incorrect predictions when based solely on 2D graph data. However, integrating SMILES information can correct these results. On the other hand, predictions based on 2D graphs for molecules like fluphenazine are accurate, whereas SMILES predictions are not. Nevertheless, combining both modalities does not compromise the overall accuracy of the final assessment. This underscores the integration of 2D graph modality data with SMILES information, which can enhance the model's ability to correct erroneous predictions and safeguard against potential interferences.

However, the integration of cross-attention might inadvertently reduce the impact of the singular modalities. Nonetheless, the strategic implementation of cross-attention promotes an effective fusion of SMILES and graph2d, resulting in combined vectors that outperform the individual modalities. As shown in Table 4, the methods involving modality alignment, which utilize contrastive learning between SMILES and 2D graphs, improve the model's performance from 65.2% (Index 1) to 72.0% (average of Indexes 4, 5, and 7). Furthermore, the methods integrating cross-attention mechanisms for modality fusion further enhance the model's performance to 74.3% (average of Indexes 9 and 12). GraphMVP, using contrastive learning between 2D and 3D graphs, elevates the performance from 65.2% to 71.7% (Indexes 1 and 2). The MolLM achieves the optimal performance of 78.4% (Index 13) through contrastive learning and the fusion of 2D and 3D graphs. This reveals that multimodal learning based on 2D graphs offers a performance increase of 6.5% (71.7%−65.2%) to 6.8% (72.0%−65.2%) over single-modality learning, with the attention fusion mechanism providing an additional 2.3% (74.3%−72.0%) to 6.7% (78.4%−71.7%) boost.

In the realm of molecular property prediction, the application of multi-modality methods introduces significant challenges and limitations. Key among these is the substantial increase in computational resource consumption required for processing complex multi-modal data. This issue is particularly evident when generating detailed 3D representations from standard molecular formats like SMILES, which demands extensive resources. Concurrently, these methods often contend with processing redundant information, as a result of overlapping content among various modalities, such as SMILES, 2D graphs, and 3D structures. This overlap leads to inefficiencies due to the repeated processing of identical molecular characteristics. Additionally, integrating different data forms into a cohesive model adds another layer of complexity, necessitating a strategic approach for effective data combination to enhance predictive accuracy.

Despite these challenges, the conclusion drawn from our findings is clear: by anchoring on 2D graphs and enriching them with 1D SMILES or 3D graph information, multi-modal learning has achieved a significant ROC-AUC uplift of 9.1% to 13.2% compared to single-modality models (results from Indexes 1, 9, 12, and 13 in Table 4). These results underscore the substantial advantages and vast potential of modality fusion techniques in providing more holistic and comprehensive insights into molecular structures, thus enhancing the overall predictive accuracy in molecular property prediction.

7 Conclusion

In this paper, we discuss the significant role of multi-modal data and domain knowledge in enhancing molecular property prediction through DL methods. We explore various molecular modalities and domain knowledge, crucial in understanding molecular complexities. Our review of recent encoder architectures and training strategies highlights how integrating domain knowledge and multi-modal data advances these models. By benchmarking prominent works, we provide a comparative analysis of their effectiveness. Ultimately, our discussion reveals the profound impact of domain knowledge and multi-modal data in DL approaches, marking a transformative advancement in drug discovery and computational molecular analysis.

References

[1] J. Shen and C. A Nicolaou, Molecular property prediction: Recent trends in the era of artificial intelligence, *Drug Discov Today Technol.*, vol. 32–33, pp. 29–36, 2019.

[2] Z. Li, M. Jiang, S. Wang, and S. Zhang, Deep learning methods for molecular representation and property prediction, *Drug Discov. Today*, vol. 27, no. 12, p. 103373, 2022.

[3] H. Ma, C. Yan, Y. Guo, S. Wang, Y. Wang, H. Sun, and J. Huang, Improving molecular property prediction on limited data with deep multi-label learning, in *Proc. 2020 IEEE Int. Conf. Bioinformatics and Biomedicine* (*BIBM*), Seoul, Republic of Korea, 2020, pp. 2779–2784.

[4] X. Lin, Z. Quan, Z. J. Wang, H. Huang, and X. Zeng, A novel molecular representation with BiGRU neural networks for learning atom, *Brief. Bioinform.*, vol. 21, no. 6, pp. 2099–2111, 2020.

[5] Q. Lv, G. Chen, L. Zhao, W. Zhong, and C. Y. C. Chen, Mol2Context-vec: Learning molecular representation from context awareness for drug discovery, *Brief. Bioinform.*, vol. 22, no. 6, p. bbab317, 2021.

[6] S. Han, H. Fu, Y. Wu, G. Zhao, Z. Song, F. Huang, Z. Zhang, S. Liu, and W. Zhang, HimGNN: A novel hierarchical molecular graph representation learning framework for property prediction, *Brief. Bioinform.*, vol.

24, no. 5, p. bbad305, 2023.

[7] G. Bouritsas, F. Frasca, S. Zafeiriou, and M. M. Bronstein, Improving graph neural network expressivity via subgraph isomorphism counting, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 657–668, 2023.

[8] Y. Song, S. Zheng, Z. Niu, Z. H. Fu, Y. Lu, and Y. Yang, Communicative representation learning on attributed molecular graphs, in *Proc. 29th Int. Joint Conf. Artificial Intelligence*, Yokohama, Japan, 2020, pp. 2831–2838.

[9] H. Li, D. Zhao, and J. Zeng, KPGT: Knowledge-guided pre-training of graph transformer for molecular property prediction, in *Proc. 28th ACM SIGKDD Conf. Knowledge Discovery and Data Mining*, Washington, DC, USA, 2022, pp. 857–867.

[10] Z. Xiong, D. Wang, X. Liu, F. Zhong, X. Wan, X. Li, Z. Li, X. Luo, K. Chen, H. Jiang, et al., Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism, *J. Med. Chem.*, vol. 63, no. 16, pp. 8749–8760, 2020.

[11] Y. Rong, Y. Bian, T. Xu, W. Xie, Y. Wei, W. Huang, and J. Huang, Self-supervised graph transformer on large-scale molecular data, in *Proc. 34th Int. Conf. Neural Information Processing Systems*, Vancouver, Canada, 2020, pp. 12559–12571.

[12] J. Ross, B. Belgodere, V. Chenthamarakshan, I. Padhi, Y. Mroueh, and P. Das, Large-scale chemical language representations capture molecular structure and properties, *Nat. Mach. Intell.*, vol. 4, no. 12, pp. 1256–1264, 2022.

[13] S. Yin and G. Zhong, LGI-GT: Graph transformers with local and global operators interleaving, in *Proc. 32nd Int. Joint Conf. Artificial Intelligence*, Macao, China, 2023, pp. 4504–4512.

[14] S. Luo, T. Chen, Y. Xu, S. Zheng, T. Y. Liu, L. Wang, and D. He, One transformer can understand both 2D & 3D molecular data, arXiv preprint arXiv: 2210.01765, 2023.

[15] X. Zeng, H. Xiang, L. Yu, J. Wang, K. Li, R. Nussinov, and F. Cheng, Accurate prediction of molecular properties and drug targets using a self-supervised image representation learning framework, *Nat. Mach. Intell.*, vol. 4, no. 11, pp. 1004–1016, 2022.

[16] J. H. Chen and Y. J. Tseng, Different molecular enumeration influences in deep learning: An example using aqueous solubility, *Brief. Bioinform.*, vol. 22, no. 3, p. bbaa092, 2021.

[17] S. Liu, J. Li, K. C. Bennett, B. Ganoe, T. Stauch, M. Head-Gordon, A. Hexemer, D. Ushizima, and T. Head-Gordon, Multiresolution 3D-DenseNet for chemical shift prediction in NMR crystallography, *J. Phys. Chem. Lett.*, vol. 10, no. 16, p. 4558–4565, 2019.

[18] S. Liu, H. Wang, W. Liu, J. Lasenby, H. Guo, and J. Tang, Pre-training molecular graph representation with 3D geometry, arXiv preprint arXiv: 2110.07728, 2022.

[19] S. Li, J. Zhou, T. Xu, D. Dou, and H. Xiong, GeomGCL: Geometric graph contrastive learning for molecular property prediction, in *Proc. 36th AAAI Conf. Artificial Intelligence*, Virtual Event, 2022, pp. 4541–4549.

[20] J. Zhu, Y. Xia, L. Wu, S. Xie, T. Qin, W. Zhou, H. Li, and T. Y. Liu, Unified 2D and 3D pre-training of molecular representations, in *Proc. 28th ACM SIGKDD Conf. Knowledge Discovery and Data Mining*, Washington, DC, USA, 2022, pp. 2626–2636.

[21] Z. Guo, W. Yu, C. Zhang, M. Jiang, and N. V. Chawla, GraSeq: Graph and sequence fusion learning for molecular property prediction, in *Proc. 29th ACM Int. Conf. Information & Knowledge Management*, Virtual Event, 2020, pp. 435–443.

[22] Y. Wang, J. Wang, Z. Cao, and A. B. Farimani, Molecular contrastive learning of representations via graph neural networks, *Nat. Mach. Intell.*, vol. 4, no. 3, pp. 279–287, 2022.

[23] Y. Fang, Q. Zhang, N. Zhang, Z. Chen, X. Zhuang, X. Shao, X. Fan, and H. Chen, Knowledge graph-enhanced molecular contrastive learning with functional prompt, *Nat. Mach. Intell.*, vol. 5, no. 5, pp. 542–553, 2023.

[24] H. Li, R. Zhang, Y. Min, D. Ma, D. Zhao, and J. Zeng, A knowledge-guided pre-training framework for improving molecular representation learning, *Nat. Commun.*, vol. 14, no. 1, p. 7568, 2023.

[25] Z. Hao, C. Lu, Z. Huang, H. Wang, Z. Hu, Q. Liu, E. Chen, and C. Lee, ASGN: An active semi-supervised graph neural network for molecular property prediction, in *Proc. 26th ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining*, Virtual Event, 2020, pp. 731–752.

[26] F. Y. Sun, J. Hoffmann, V. Verma, and J. Tang, InfoGraph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization, arXiv preprint arXiv: 1908.01000, 2020.

[27] D. Zhang, W. Feng, Y. Wang, Z. Qi, Y. Shan, and J. Tang, DropConn: Dropout connection based random GNNs for molecular property prediction, *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 2, pp. 518–529, 2024.

[28] Y. Sun, Y. Chen, W. Ma, W. Huang, K. Liu, Z. Ma, W. Y. Ma, and Y. Lan, PEMP: Leveraging physics properties to enhance molecular property prediction, in *Proc. 31st ACM Int. Conf. Information & Knowledge Management*, Atlanta, GA, USA, 2022, pp. 3505–3513.

[29] W. Chen, A. Tripp, and J. M. Hernández-Lobato, Meta-learning adaptive deep kernel Gaussian processes for molecular property prediction, arXiv preprint arXiv: 2205.02708, 2023.

[30] X. Zhuang, Q. Zhang, B. Wu, K. Ding, Y. Fang, and H. Chen, Graph sampling-based meta-learning for molecular property prediction, arXiv preprint arXiv: 2306.16780, 2023.

[31] S. Biswas, Y. Chung, J. Ramirez, H. Wu, and W. H. Green, Predicting critical properties and acentric factors of fluids using multitask machine learning, *J. Chem. Inf. Model.*, vol. 63, no. 15, pp. 4574–4588, 2023.

[32] Z. Tan, Y. Li, W. Shi, and S. Yang, A multitask approach

to learn molecular properties, *J. Chem. Inf. Model.*, vol. 61, no. 8, pp. 3824–3834, 2021.

[33] 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, *Chem. Sci.*, vol. 9, no. 2, pp. 513–530, 2018.

[34] D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, *J. Chem. Inf. Comput. Sci.*, vol. 28, no. 1, pp. 31–36, 1988.

[35] D. Weininger, A. Weininger, and J. L. Weininger, SMILES. 2. Algorithm for generation of unique smiles notation, *J. Chem. Inf. Comput. Sci*., vol. 29, no. 2, pp. 97–101, 1989.

[36] D. Weininger, SMILES. 3. DEPICT. Graphical depiction of chemical structures, *J. Chem. Inf. Comput. Sci*., vol. 30, no. 3, pp. 237–243, 1990.

[37] D. Rogers and M. Hahn, Extended-connectivity fingerprints, *J. Chem. Inf. Model.*, vol. 50, no. 5, pp. 742–754, 2010.

[38] J. L. Durant, B. A. Leland, D. R. Henry, and J. G. Nourse, Reoptimization of mdl keys for use in drug discovery, *J. Chem. Inf. Comput. Sci.*, vol. 42, no. 6, pp. 1273–1280, 2002.

[39] M. Krenn, F. Häse, A. Nigam, P. Friederich, and A. Aspuru-Guzik, Self-referencing embedded strings (SELFIES): A 100% robust molecular string representation, *Mach. Learn.*: *Sci. Technol.*, vol. 1, no. 4, p. 045024, 2020.

[40] A. D. McNaught and A. Wilkinson, *Compendium of Chemical Terminology*, 2nd ed. Oxford, UK: Blackwell Science, 1997.

[41] S. R. Heller, A. McNaught, I. Pletnev, S. Stein, and D. Tchekhovskoi, InChi, the IUPAC international chemical identifier, *J. Cheminform.*, vol. 7, no. 1, p. 23, 2015.

[42] G. Landrum, RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling, *Greg Landrum*, vol. 8, p. 31, 2013.

[43] W. L. DeLano, PyMOL: An open-source molecular graphics tool, *CCP4 Newsl. Protein Crystallogr*, vol. 40, no. 1, pp. 82–92, 2002.

[44] J. Sunseri and D. R. Koes, Libmolgrid: Graphics processing unit accelerated molecular gridding for deep learning applications, *J. Chem. Inf. Model.*, vol. 60, no. 3, pp. 1079–1084, 2020.

[45] J. Degen, C. Wegscheid-Gerlach, A. Zaliani, and M. Rarey, On the art of compiling and using 'drug-like' chemical fragment spaces, *ChemMedChem*, vol. 3, no. 10, pp. 1503–1507, 2008.

[46] X. Q. Lewell, D. B. Judd, S. P. Watson, and M. M. Hann, RECAP-retrosynthetic combinatorial analysis procedure: A powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry, *J. Chem. Inf. Comput. Sci*., vol. 38, no. 3, pp. 511–522, 1998.

[47] G. W. Bemis and M. A. Murcko, The properties of known drugs. 1. Molecular frameworks, *J. Med. Chem.*, vol. 39, no. 15, pp. 2887–2893, 1996.

[48] T. Liu, M. Naderi, C. Alvin, S. Mukhopadhyay, and M. Brylinski, Break down in order to build up: Decomposing small molecules for fragment-based drug design with *e*MolFrag, *J. Chem. Inf. Model.*, vol. 57, no. 4, pp. 627–631, 2017.

[49] F. Kruger, N. Stiefl, and G. A. Landrum, rdScaffoldNetwork: The scaffold network implementation in RDKit, *J. Chem. Inf. Model*., vol. 60, no. 7, pp. 3331–3335, 2020.

[50] C. K. Wu, X. C. Zhang, Z. J. Yang, A. P. Lu, T. J. Hou, and D. S. Cao, Learning to SMILES: Ban-based strategies to improve latent representation learning from molecules, *Brief. Bioinform.*, vol. 22, no. 6, p. bbab327, 2021.

[51] K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea, et al., Analyzing learned molecular representations for property prediction, *J. Chem. Inf. Model.*, vol. 59, no. 8, pp. 3370–3388, 2019.

[52] Y. Ji, G. Wan, Y. Zhan, and B. Du, Metapath-fused heterogeneous graph network for molecular property prediction, *Inf. Sci.*, vol. 629, pp. 155–168, 2023.

[53] X. Fang, L. Liu, J. Lei, D. He, S. Zhang, J. Zhou, F. Wang, H. Wu, and H. Wang, Geometry-enhanced molecular representation learning for property prediction, *Nat. Mach. Intell.*, vol. 4, no. 2, pp. 127–134, 2022.

[54] G. Zhou, Z. Gao, Q. Ding, H. Zheng, H. Xu, Z. Wei, L. Zhang, and G. Ke, Uni-Mol: A universal 3D molecular representation learning framework, chemRxiv, doi: 10.26434/chemrxiv-2022-jjm0j-v4.

[55] S. Chithrananda, G. Grand, and B. Ramsundar, ChemBERTa: Large-scale self-supervised pretraining for molecular property prediction, arXiv preprint arXiv: 2010.09885, 2020.

[56] A. Yüksel, E. Ulusoy, A. Ünlü, and T. Doğan, SELFormer: Molecular representation learning via SELFIES language models, *Mach. Learn.*: *Sci. Technol*., vol. 4, no. 2, p. 025035, 2023.

[57] X. C. Zhang, J. C. Yi, G. P. Yang, C. K. Wu, T. J. Hou, and D. S. Cao, ABC-Net: A divide-and-conquer based deep learning architecture for SMILES recognition from molecular images, *Brief. Bioinform*., vol. 23, no. 2, p. bbac033, 2022.

[58] S. Liu, W. Nie, C. Wang, J. Lu, Z. Qiao, L. Liu, J. Tang, C. Xiao, and A. Anandkumar, Multi-modal molecule structure–text model for text-based retrieval and editing, *Nat. Mach. Intell.*, vol. 5, no. 12, pp. 1447–1457, 2023.

[59] P. Liu, X. Qiu, X. Chen, S. Wu, and X. Huang, Multi-timescale long short-term memory neural network for modelling sentences and documents, in *Proc. 2015 Conf. Empirical Methods in Natural Language Processing*, Lisbon, Portugal, 2015, pp. 2326–2335.

[60] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, Gated feedback recurrent neural networks, in *Proc. 32nd Int.*

Conf. Int. Conf. Machine Learning, Lille, France, 2015, pp. 2067–2075.

[61] A. L. Nazarova, L. Yang, K. Liu, A. Mishra, R. K. Kalia, K. I. Nomura, A. Nakano, P. Vashishta, and P. Rajak, Dielectric polymer property prediction using recurrent neural networks with optimizations, *J. Chem. Inf. Model.*, vol. 61, no. 5, pp. 2175–2186, 2021.

[62] Z. Wang, Y. Su, W. Shen, S. Jin, J. H. Clark, J. Ren, and X. Zhang, Predictive deep learning models for environmental properties: The direct calculation of octanol–water partition coefficients from molecular graphs, *Green Chem.*, vol. 21, no. 16, pp. 4555–4565, 2019.

[63] M. Withnall, E. Lindelöf, O. Engkvist, and H. Chen, Building attention and edge message passing neural networks for bioactivity and physical–chemical property prediction, *J. Cheminform.*, vol. 12, no. 1, p. 1, 2020.

[64] P. Li, Y. Li, C. Y. Hsieh, S. Zhang, X. Liu, H. Liu, S. Song, and X. Yao, TrimNet: Learning molecular representation from triplet messages for biomedicine, *Brief. Bioinform.*, vol. 22, no. 4, p. bbaa266, 2021.

[65] X. Zhang, C. Chen, Z. Meng, Z. Yang, H. Jiang, and X. Cui, CoAtGIN: Marrying convolution and attention for graph-based molecule property prediction, in *Proc. 2022 IEEE Int. Conf. Bioinformatics and Biomedicine* (*BIBM*), Las Vegas, NV, USA, 2022, pp. 374–379.

[66] X. Fan, M. Gong, Y. Wu, A. K. Qin, and Y. Xie, Propagation enhanced neural message passing for graph representation learning, *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 2, pp. 1952–1964, 2023.

[67] Y. Li, P. Li, X. Yang, C. Y. Hsieh, S. Zhang, X. Wang, R. Lu, H. Liu, and X. Yao, Introducing block design in graph neural networks for molecular properties prediction, *Chem. Eng. J.*, vol. 414, p. 128817, 2021.

[68] H. Ma, Y. Bian, Y. Rong, W. Huang, T. Xu, W. Xie, G. Ye, and J. Huang, Multi-view graph neural networks for molecular property prediction, arXiv preprint arXiv: 2005.13607, 2020.

[69] X. Liu, X. Wang, J. Wu, and K. Xia, Hypergraph-based persistent cohomology (HPC) for molecular representations in drug design, *Brief. Bioinform.*, vol. 22, no. 5, p. bbaa411, 2021.

[70] J. Feng, Z. Wang, Y. Li, B. Ding, Z. Wei, and H. Xu, MGMAE: Molecular representation learning by reconstructing heterogeneous graphs with a high mask ratio, in *Proc. 31st ACM Int. Conf. Information & Knowledge Management*, Atlanta, GA, USA, 2022, pp. 509–519.

[71] T. Hasebe, Knowledge-embedded message-passing neural networks: Improving molecular property prediction with human knowledge, *ACS Omega*, vol. 6, no. 42, pp. 27955–27967, 2021.

[72] S. Yang, Z. Li, G. Song, and L. Cai, Deep molecular representation learning via fusing physical and chemical information, in *Proc. Annu. Conf. Neural Information Processing Systems*, Virtual Event, 2021, pp. 16346–16357.

[73] X. Zang, X. Zhao, and B. Tang, Hierarchical molecular graph self-supervised learning for property prediction, *Commun. Chem.*, vol. 6, no. 1, p. 34, 2023.

[74] N. Liu, S. Jian, D. Li, Y. Zhang, Z. Lai, and H. Xu, Hierarchical adaptive pooling by capturing high-order dependency for graph representation learning, *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 3952–3965, 2023.

[75] J. Gao, J. Gao, X. Ying, M. Lu, and J. Wang, Higher-order interaction goes neural: A substructure assembling graph attention network for graph classification, *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 2, pp. 1594–1608, 2023.

[76] X. B. Ye, Q. Guan, W. Luo, L. Fang, Z. R. Lai, and J. Wang, Molecular substructure graph attention network for molecular property identification in drug discovery, *Pattern Recog.*, vol. 128, p. 108659, 2022.

[77] W. Zhu, Y. Zhang, D. Zhao, J. Xu, and L. Wang, HiGNN: A hierarchical informative graph neural network for molecular property prediction equipped with feature-wise attention, *J. Chem. Inf. Model.*, vol. 63, no. 1, pp. 43–55, 2023.

[78] C. Lu, Q. Liu, C. Wang, Z. Huang, P. Lin, and L. He, Molecular property prediction: A multilevel quantum interactions modeling perspective, in *Proc. 33rd AAAI Conf. Artificial Intelligence*, Honolulu, HI, USA, 2019, pp. 1052–1060.

[79] M. Fey, J. G. Yuen, and F. Weichert, Hierarchical inter-message passing for learning on molecular graphs, arXiv preprint arXiv: 2006.12179, 2020.

[80] F. Wu, D. Radev, and S. Z. Li, Molformer: Motif-based transformer on 3D heterogeneous molecular graphs, in *Proc. 37th AAAI Conf. Artificial Intelligence*, Washington, DC, USA, 2023, pp. 5312–5320.

[81] F. B. Fuchs, D. E. Worrall, V. Fischer, and M. Welling, SE(3)-transformers: 3D roto-translation equivariant attention networks, arXiv preprint arXiv: 2006.10503, 2020.

[82] K. T. Schütt, O. T. Unke, and M. Gastegger, Equivariant message passing for the prediction of tensorial properties and molecular spectra, arXiv preprint arXiv: 2102.03150, 2021.

[83] J. Brandstetter, R. Hesselink, E. van der Pol, E. J. Bekkers, and M. Welling, Geometric and physical quantities improve E(3) equivariant message passing, arXiv preprint arXiv: 2110.02905, 2022.

[84] J. Gasteiger, F. Becker, and S. Günnemann, GemNet: Universal directional graph neural networks for molecules, arXiv preprint arXiv: 2106.08903, 2022.

[85] J. Gasteiger, S. Giri, J. T. Margraf, and S. Günnemann, Fast and uncertainty-aware directional message passing for non-equilibrium molecules, arXiv preprint arXiv: 2011.14115, 2022.

[86] M. Shuaibi, A. Kolluru, A. Das, A. Grover, A. Sriram, Z. Ulissi, and C. L. Zitnick, Rotation invariant graph neural

networks using spin convolutions, arXiv preprint arXiv: 2106.09575, 2021.

[87] S. Wang, Y. Guo, Y. Wang, H. Sun, and J. Huang, SMILES-BERT: Large scale unsupervised pre-training for molecular property prediction, in *Proc. 10th ACM Int. Conf. Bioinformatics, Computational Biology and Health Informatics*, Niagara Falls, NY, USA, 2019, pp. 429–436.

[88] Y. Wang, X. Chen, Y. Min, and J. Wu, MolCloze: A unified cloze-style self-supervised molecular structure learning model for chemical property prediction, in *Proc. 2021 IEEE Int. Conf. Bioinformatics and Biomedicine* (*BIBM*), Houston, TX, USA, 2021, pp. 2896–2903.

[89] B. Winter, C. Winter, J. Schilling, and A. Bardow, A smile is all you need: Predicting limiting activity coefficients from smiles with natural language processing, *Digit Discov*, vol. 1, no. 6, pp. 859–869, 2022.

[90] J. Su, M. Ahmed, Y. Lu, S. Pan, W. Bo, and Y. Liu, RoFormer: Enhanced transformer with rotary position embedding, *Neurocomputing*, vol. 568, p. 127063, 2024.

[91] Ł. Maziarka, T. Danel, S. Mucha, K. Rataj, J. Tabor, and S. Jastrzebski, Molecule attention transformer, arXiv preprint arXiv: 2002.08264, 2020.

[92] W. Park, W. Chang, D. Lee, J. Kim, and S. W. Hwang, GRPE: Relative positional encoding for graph transformer, arXiv preprint arXiv: 2201.12787, 2022.

[93] M. S. Hussain, M. J. Zaki, and D. Subramanian, Global self-attention as a replacement for graph convolution, in *Proc. 28th ACM SIGKDD Conf. Knowledge Discovery and Data Mining*, Washington, DC, USA, 2022, pp. 655–665.

[94] D. Masters, J. Dean, K. Klaser, Z. Li, S. Maddrell-Mander, A. Sanders, H. Helal, D. Beker, L. Rampášek, and D. Beaini, GPS++: An optimised hybrid MPNN/transformer for molecular property prediction, arXiv preprint arXiv: 2212.02229, 2022.

[95] Z. Chen, H. Tan, T. Wang, T. Shen, T. Lu, Q. Peng, C. Cheng, and Y. Qi, Graph propagation transformer for graph representation learning, arXiv preprint arXiv: 2305.11424, 2023.

[96] G. P. Ren, K. J. Wu, and Y. He, Enhancing molecular representations via graph transformation layers, *J. Chem. Inf. Model.*, vol. 63, no. 9, pp. 2679–2688, 2023.

[97] J. Gao, Z. Shen, Y. Xie, J. Lu, Y. Lu, S. Chen, Q. Bian, Y. Guo, L. Shen, J. Wu, et al., TransFoxMol: Predicting molecular property with focused attention, *Brief. Bioinform.*, vol. 24, no. 5, p. bbad306, 2023.

[98] Y. Jiang, S. Jin, X. Jin, X. Xiao, W. Wu, X. Liu, Q. Zhang, X. Zeng, G. Yang, and Z. Niu, Pharmacophoric-constrained heterogeneous graph transformer model for molecular property prediction, *Commun. Chem.*, vol. 6, no. 1, p. 60, 2023.

[99] M. Hirohara, Y. Saito, Y. Koda, K. Sato, and Y. Sakakibara, Convolutional neural network based on smiles representation of compounds for detecting chemical motif, *BMC Bioinformatics*, vol. 19, no. S19, p. 526, 2018.

[100] P. Jiang, Y. Chi, X. S. Li, Z. Meng, X. Liu, X. S. Hua, and K. Xia, Molecular persistent spectral image (Mol-PSI) representation for machine learning models in drug design, *Brief. Bioinform.*, vol. 23, no. 1, p. bbab527, 2022.

[101] D. Kuzminykh, D. Polykovskiy, A. Kadurin, A. Zhebrak, I. Baskov, S. Nikolenko, R. Shayakhmetov, and A. Zhavoronkov, 3D molecular representations based on the wave transform for convolutional neural networks, *Mol. Pharm.*, vol. 15, no. 10, pp. 4378–4385, 2018.

[102] H. Cai, H. Zhang, D. Zhao, J. Wu, and L. Wang, FP-GNN: A versatile deep learning architecture for enhanced molecular property prediction, *Brief. Bioinform.*, vol. 23, no. 6, p. bbac408, 2022.

[103] X. Wang, Z. Li, M. Jiang, S. Wang, S. Zhang, and Z. Wei, Molecule property prediction based on spatial graph embedding, *J. Chem. Inf. Model.*, vol. 59, no. 9, pp. 3817–3828, 2019.

[104] J. Liu, X. Lei, Y. Zhang, and Y. Pan, The prediction of molecular toxicity based on BiGRU and GraphSAGE, *Comput. Biol. Med.*, vol. 153, p. 106524, 2023.

[105] Y. Luo, K. Yang, M. Hong, X. Liu, and Z. Nie, MolFM: A multimodal molecular foundation model, arXiv preprint arXiv: 2307.09484, 2023.

[106] Y. Sun, M. Islam, E. Zahedi, M. Kuenemann, H. Chouaib, and P. Hu, Molecular property prediction based on bimodal supervised contrastive learning, in *Proc. 2022 IEEE Int. Conf. Bioinformatics and Biomedicine* (*BIBM*), Las Vegas, NV, USA, 2022, pp. 394–397.

[107] P. Liu, Y. Ren, J. Tao, and Z. Ren, GIT-Mol: A multi-modal large language model for molecular science with graph, image, and text, arXiv preprint arXiv: 2308.06911, 2024.

[108] Q. Tang, F. Nie, Q. Zhao, and W. Chen, A merged molecular representation deep learning method for blood–brain barrier permeability prediction, *Brief. Bioinform.*, vol. 23, no. 5, p. bbac357, 2022.

[109] T. Zhang, S. Chen, A. Wulamu, X. Guo, Q. Li, and H. Zheng, TransG-Net: Transformer and graph neural network based multi-modal data fusion network for molecular properties prediction, *Appl. Intell.*, vol. 53, no. 12, pp. 16077–16088, 2023.

[110] D. Chen, K. Gao, D. D. Nguyen, X. Chen, Y. Jiang, G. W. Wei, and F. Pan, Algebraic graph-assisted bidirectional transformers for molecular property prediction, *Nat. Commun.*, vol. 12, no. 1, p. 3521, 2021.

[111] W. X. Shen, X. Zeng, F. Zhu, Y. L. Wang, C. Qin, Y. Tan, Y. Y. Jiang, and Y. Z. Chen, Out-of-the-box deep learning prediction of pharmaceutical properties by broadly learned knowledge-based molecular representations, *Nat. Mach. Intell.*, vol. 3, no. 4, pp. 334–343, 2021.

[112] Y. Liu, L. Wang, M. Liu, X. Zhang, B. Oztekin, and S. Ji, Spherical message passing for 3D molecular graphs, arXiv preprint arXiv: 2102.05013, 2022.

[113] Z. Wang, M. Liu, Y. Luo, Z. Xu, Y. Xie, L. Wang, L. Cai, Q. Qi, Z. Yuan, T. Yang, et al., Advanced graph and sequence neural networks for molecular property prediction and drug discovery, *Bioinformatics*, vol. 38, no. 9, pp. 2579–2586, 2022.

[114] J. Zhu, Y. Xia, L. Wu, S. Xie, W. Zhou, T. Qin, H. Li, and T. Y. Liu, Dual-view molecular pre-training, in *Proc. 29th ACM SIGKDD Conf. Knowledge Discovery and Data Mining*, Long Beach, CA, USA, 2023, pp. 3615–3627.

[115] M. Sun, J. Xing, H. Wang, B. Chen, and J. Zhou, MoCL: Data-driven molecular fingerprint via knowledge-aware contrastive learning from molecular graph, in *Proc. 27th ACM SIGKDD Conf. Knowledge Discovery & Data Mining*, Singapore, 2021, p. 3585–3594.

[116] Y. Wang, R. Magar, C. Liang, and A. B. Farimani, Improving molecular contrastive learning via faulty negative mitigation and decomposed fragment contrast, *J. Chem. Inf. Model.*, vol. 62, no. 11, pp. 2713–2725, 2022.

[117] Y. You, T. Chen, Y. Sui, T. Chen, Z. Wang, and Y. Shen, Graph contrastive learning with augmentations, in *Proc. 34th Int. Conf. Neural Information Processing Systems*, Vancouver, Canada, 2020, pp. 5812–5823.

[118] J. Xia, C. Zhao, B. Hu, Z. Gao, C. Tan, Y. Liu, S. Li, and S. Z. Li, Mole-BERT: Rethinking pre-training graph neural networks for molecules, chemRxiv, doi: 10.26434/chemrxiv-2023-dngg4.

[119] Z. Wu, D. Jiang, J. Wang, X. Zhang, H. Du, L. Pan, C. Y. Hsieh, D. Cao, and T. Hou, Knowledge-based BERT: A method to extract molecular features like computational chemists, *Brief. Bioinform.*, vol. 23, no. 3, p. bbac131, 2022.

[120] Z. Wu, J. Wang, H. Du, D. Jiang, Y. Kang, D. Li, P. Pan, Y. Deng, D. Cao, C. Y. Hsieh, et al., Chemistry-intuitive explanation of graph neural networks for molecular property prediction with substructure masking, *Nat. Commun.*, vol. 14, no. 1, p. 2585, 2023.

[121] S. Kim, J. Nam, J. Kim, H. Lee, S. Ahn, and J. Shin, Fragment-based multi-view molecular contrastive learning, in *Proc. ICLR 2023*, https://openreview.net/forum?id=9lGwd4q8KJc, 2024.

[122] F. Wu, H. Qin, S. Li, S. Z. Li, X. Zhan, and J. Xu, InstructBio: A large-scale semi-supervised learning paradigm for biochemical problems, arXiv preprint arXiv: 2304.03906, 2023.

[123] Q. Lv, G. Chen, Z. Yang, W. Zhong, and C. Y. C. Chen, Meta learning with graph attention networks for low-data drug discovery, *IEEE Trans. Neural Netw. Learn. Syst.*, doi: 10.1109/TNNLS.2023.3250324.

[124] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in *Proc. 2019 Conf. North American Chapter of the Association for Computational Linguistics*: *Human Language Technologies, Volume 1* (*Long and Short Papers*), Minneapolis, MN, USA, 2019, pp. 4171–4186.

[125] L. Floridi and M. Chiriatti, GPT-3: Its nature, scope, limits, and consequences, *Minds Mach.*, vol. 30, pp. 681–694, 2020.

[126] X. C. Zhang, C. K. Wu, Z. J. Yang, Z. X. Wu, J. C. Yi, C. Y. Hsieh, T. J. Hou, and D. S. Cao, MG-BERT: Leveraging unsupervised atomic representation learning for molecular property prediction, *Brief. Bioinform.*, vol. 22, no. 6, p. bbab152, 2021.

[127] W. Ahmad, E. Simon, S. Chithrananda, G. Grand, and B. Ramsundar, ChemBERTa-2: Towards chemical foundation models, arXiv preprint arXiv: 2209.01712, 2022.

[128] R. Irwin, S. Dimitriadis, J. He, and E. J. Bjerrum, Chemformer: A pre-trained transformer for computational chemistry, *Mach. Learn.*: *Sci. Technol.*, vol. 3, no. 1, p. 015022, 2022.

[129] W. Hu, B. Liu, J. Gomes, M. Zitnik, P. Liang, V. Pande, and J. Leskovec, Strategies for pre-training graph neural networks, arXiv preprint arXiv: 1905.12265, 2020.

[130] J. Godwin, M. Schaarschmidt, A. Gaunt, A. Sanchez-Gonzalez, Y. Rubanova, P. Veličković, J. Kirkpatrick, and P. Battaglia, Simple GNN regularisation for 3D molecular property prediction & beyond, arXiv preprint arXiv: 2106.07971, 2022.

[131] S. Liu, H. Guo, and J. Tang, Molecular geometry pretraining with SE(3)-invariant denoising distance matching, arXiv preprint arXiv: 2206.13602, 2023.

[132] S. Feng, Y. Ni, Y. Lan, Z. M. Ma, and W. Y. Ma, Fractional denoising for 3D molecular pre-training, in *Proc. 40th Int. Conf. Machine Learning*, Honolulu, HI, USA, 2023, pp. 9938–9961.

[133] R. Jiao, J. Han, W. Huang, Y. Rong, and Y. Liu, Energy-motivated equivariant pretraining for 3D molecular graphs, in *Proc. 37th AAAI Conf. Artificial Intelligence*, Washington, DC, USA, 2023, pp. 8096–8104.

[134] X. Gao, W. Gao, W. Xiao, Z. Wang, C. Wang, and L. Xiang, Supervised pretraining for molecular force fields and properties prediction, arXiv preprint arXiv: 2211.14429, 2022.

[135] X. Wang, H. Zhao, W. W. Tu, and Q. Yao, Automated 3D pre-training for molecular property prediction, in *Proc. 29th ACM SIGKDD Conf. Knowledge Discovery and Data Mining*, Long Beach, CA, USA, 2023, pp. 2419–2430.

[136] L. Zeng, L. Li, and J. Li, MolKD: Distilling cross-modal knowledge in chemical reactions for molecular property prediction, arXiv preprint arXiv: 2305.01912, 2023.

[137] J. Broberg, M. Bånkestad, and E. Ylipää, Pre-training transformers for molecular property prediction using reaction prediction, arXiv preprint arXiv: 2207.02724, 2022.

[138] X. C. Zhang, C. K. Wu, J. C. Yi, X. X. Zeng, C. Q. Yang, A. P. Lu, T. J. Hou, and D. S. Cao, Pushing the boundaries of molecular property prediction for drug discovery with multitask learning BERT enhanced by SMILES enumeration, *Research*, vol. 2022, p. 0004,

2022.

[139] H. Abdel-Aty and I. R. Gould, Large-scale distributed training of transformers for chemical fingerprinting, *J. Chem. Inf. Model.*, vol. 62, no. 20, pp. 4852–4862, 2022.

[140] Z. Zheng, Y. Tan, H. Wang, S. Yu, T. Liu, and C. Liang, CasANGCL: Pre-training and fine-tuning model based on cascaded attention network and graph contrastive learning for molecular property prediction, *Brief. Bioinform.*, vol. 24, no. 1, p. bbac566, 2023.

[141] X. Guan and D. Zhang, T-MGCL: Molecule graph contrastive learning based on transformer for molecular property prediction, *IEEE/ACM Trans. Comput. Biol. Bioinform*., vol. 20, no. 6, pp. 3851–3862, 2023.

[142] H. Liu, Y. Huang, X. Liu, and L. Deng, Attention-wise masked graph contrastive learning for predicting molecular property, *Brief. Bioinform.*, vol. 23, no. 5, p. bbac303, 2022.

[143] S. Lin, C. Liu, P. Zhou, Z. Y. Hu, S. Wang, R. Zhao, Y. Zheng, L. Lin, E. Xing, and X. Liang, Prototypical graph contrastive learning, *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 2, pp. 2747–2758, 2024.

[144] J. Cui, H. Chai, Y. Gong, Y. Ding, Z. Hua, C. Gao, and Q. Liao, MocGCL: Molecular graph contrastive learning via negative selection, in *Proc. 2023 Int. Joint Conf. Neural Networks* (*IJCNN*), Gold Coast, Australia, 2023, pp. 1–8.

[145] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, Momentum contrast for unsupervised visual representation learning, in *Proc. 2020 IEEE/CVF Conf. Computer Vision and Pattern Recognition*, Seattle, WA, USA, 2020, pp. 9726–9735.

[146] M. J. Zaki and W. Meira Jr, *Data Mining and Analysis*: *Fundamental Concepts and Algorithms*. Cambridge, UK: Cambridge University Press, 2014.

[147] Y. Wang, Y. Min, E. Shao, and J. Wu, Molecular graph contrastive learning with parameterized explainable augmentations, in *Proc. 2021 IEEE Int. Conf. Bioinformatics and Biomedicine* (*BIBM*), Houston, TX, USA, 2021, pp. 1558–1563.

[148] M. Liu, Y. Yang, X. Gong, L. Liu, and Q. Liu, HierMRL: Hierarchical structure-aware molecular representation learning for property prediction, in *Proc. 2022 IEEE Int. Conf. Bioinformatics and Biomedicine* (*BIBM*), Las Vegas, NV, USA, 2022, pp. 386–389.

[149] J. Wang, J. Guan, and S. Zhou, Molecular property prediction by contrastive learning with attention-guided positive sample selection, *Bioinformatics*, vol. 39, no. 5, p. btad258, 2023.

[150] K. Moon, H. J. Im, and S. Kwon, 3D graph contrastive learning for molecular property prediction, *Bioinformatics*, vol. 39, no. 6, p. btad371, 2023.

[151] T. Kuang, Y. Ren, and Z. Ren, 3D-mol: A novel contrastive learning framework for molecular property prediction with 3D information, arXiv preprint arXiv: 2309.17366, 2024.

[152] X. Wu, J. Duan, Y. Pan, and M. Li, Medical knowledge graph: Data sources, construction, reasoning, and applications, *Big Data Mining and Analytics*, vol. 6, no. 2, pp. 201–217, 2023.

[153] R. Hua, X. Wang, C. Cheng, Q. Zhu, and X. Zhou, A chemical domain knowledge-aware framework for multi-view molecular property prediction, in *Proc. 7th China Conf. Knowledge Graph and Semantic Computing Evaluations*, Qinhuangdao, China, 2022, pp. 1–11.

[154] Y. Fang, Q. Zhang, H. Yang, X. Zhuang, S. Deng, W. Zhang, M. Qin, Z. Chen, X. Fan, and H. Chen, Molecular contrastive learning with chemical element knowledge graph, in *Proc. 36th AAAI Conf. Artificial Intelligence*, Virtual Event, 2022, pp. 3968–3976.

[155] M. Xu, H. Wang, B. Ni, H. Guo, and J. Tang, Self-supervised graph-level representation learning with local and global structure, in *Proc. 38th Int. Conf. Machine Learning*, Virtual Event, 2021, pp. 11548–11558.

[156] X. Xu, C. Deng, Y. Xie, and S. Ji, Group contrastive self-supervised learning on graphs, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3169–3180, 2023.

[157] X. Shen, Y. Liu, Y. Wu, and L. Xie, MoLGNN: Self-supervised motif learning graph neural network for drug discovery, in *Proc. Machine Learning for Molecules Workshop at NeurIPS 2020*, https://ml4molecules.github.io/papers2020/ML4Molecules_2020_paper_4.pdf, 2024.

[158] X. Luo, W. Ju, M. Qu, Y. Gu, C. Chen, M. Deng, X. S. Hua, and M. Zhang, CLEAR: Cluster-enhanced contrast for self-supervised graph representation learning, *IEEE Trans. Neural Netw. Learn. Syst*., vol. 35, no. 1, pp. 899–912, 2024.

[159] R. Benjamin, U. Singer, and K. Radinsky, Graph neural networks pretraining through inherent supervision for molecular property prediction, in *Proc. 31st ACM Int. Conf. Information & Knowledge Management*, Atlanta, GA, USA, 2022, pp. 2903–2912.

[160] G. Shi, Y. Zhu, J. K. Liu, and X. Li, Hegcl: Advance self-supervised learning in heterogeneous graph-level representation, *IEEE Trans. Neural Netw. Learn. Syst*., doi: 10.1109/TNNLS.2023.3273255.

[161] A. Xie, Z. Zhang, J. Guan, and S. Zhou, Self-supervised learning with chemistry-aware fragmentation for effective molecular property prediction, *Brief. Bioinform.*, vol. 24, no. 5, p. bbad296, 2023.

[162] Z. Ji, R. Shi, J. Lu, F. Li, and Y. Yang, ReLMole: Molecular representation learning based on two-level graph similarities, *J. Chem. Inf. Model.*, vol. 62, no. 22, pp. 5361–5372, 2022.

[163] R. Hadsell, S. Chopra, and Y. LeCun, Dimensionality reduction by learning an invariant mapping, in *Proc. 2006 IEEE Computer Society Conf. Computer Vision and Pattern Recognition* (*CVPR'06*), New York, NY, USA, 2006, pp. 1735–1742.

[164] G. A. Pinheiro, J. L. F. Da Silva, and M. G. Quiles, SMICLR: Contrastive learning on multiple molecular representations for semisupervised and unsupervised representation learning, *J. Chem. Inf. Model*., vol. 62, no.

17, pp. 3948–3960, 2022.

[165] C. Zhang, X. Yan, and Y. Liu, Pseudo-siamese neural network based graph and sequence representation learning for molecular property prediction, in *Proc. 2022 IEEE Int. Conf. Bioinformatics and Biomedicine* (*BIBM*), Las Vegas, NV, USA, 2022, pp. 3911–3913.

[166] H. Stärk, D. Beaini, G. Corso, P. Tossou, C. Dallago, S. Günnemann, and P. Liò., 3D infomax improves GNNs for molecular property prediction, in *Proc. 39th Int. Conf. Machine Learning*, Baltimore, MD, USA, 2022, pp. 20479–20502.

[167] Y. Zhu, D. Chen, Y. Du, Y. Wang, Q. Liu, and S. Wu, Molecular contrastive pretraining with collaborative featurizations, *Journal of Chemical Information and Modeling*, vol. 64, no. 4, pp. 1112–1122, 2024.

[168] A. Tarvainen and H. Valpola, Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results, in *Proc. 31st Int. Conf. Advances in Neural Information Processing Systems*, Long Beach, California, USA, 2017, pp. 1195–1204.

[169] J. Chen, Y. W. Si, C. W. Un, and S. W. I. Siu, Chemical toxicity prediction based on semi-supervised learning and graph convolutional neural network, *J. Cheminform.*, vol. 13, no. 1, p. 93, 2021.

[170] D. Berthelot, N. Carlini, I. Goodfellow, A. Oliver, N. Papernot, and C. Raffel, MixMatch: A holistic approach to semi-supervised learning, in *Proc. 33rd Int. Conf. Neural Information Processing Systems*, Vancouver, Canada, 2019, pp. 5049–5059.

[171] K. Yu, S. Visweswaran, and K. Batmanghelich, Semi-supervised hierarchical drug embedding in hyperbolic space, *J. Chem. Inf. Model.*, vol. 60, no. 12, pp. 5647–5657, 2020.

[172] H. Ma, F. Jiang, Y. Rong, Y. Guo, and J. Huang, Robust self-training strategy for various molecular biology prediction tasks, in *Proc. 13th ACM Int. Conf. Bioinformatics, Computational Biology and Health Informatics*, Northbrook, IL, USA, 2022, pp. 1–5.

[173] Z. Zhang and M. R. Sabuncu, Generalized cross entropy loss for training deep neural networks with noisy labels, in *Proc. 32nd Int. Conf. Neural Information Processing Systems*, Montréal, Canada, 2018, pp. 8792–8802.

[174] G. Liu, T. Zhao, E. Inae, T. Luo, and M. Jiang, Semi-supervised graph imbalanced regression, arXiv preprint arXiv: 2305.12087, 2023.

[175] A. R. Zamir, A. Sax, W. Shen, L. Guibas, J. Malik, and S. Savarese, Taskonomy: Disentangling task transfer learning, in *Proc. 2018 IEEE/CVF Conf. Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 3712–3722.

[176] X. Li, X. Yan, Q. Gu, H. Zhou, D. Wu, and J. Xu, DeepChemStable: Chemical stability prediction with an attention-based graph convolution network, *J. Chem. Inf. Model.*, vol. 59, no. 3, pp. 1044–1049, 2019.

[177] X. Chen and K. He, Exploring simple siamese representation learning, in *Proc. 2021 IEEE/CVF Conf. Computer Vision and Pattern Recognition*, Nashville, TN, USA, 2021, pp. 15745–15753.

[178] H. Li, X. Zhao, S. Li, F. Wan, D. Zhao, and J. Zeng, Improving molecular property prediction through a task similarity enhanced transfer learning strategy, *iScience*, vol. 25, no. 10, p. 105231, 2022.

[179] W. Ju, Z. Liu, Y. Qin, B. Feng, C. Wang, Z. Guo, X. Luo, and M. Zhang, Few-shot molecular property prediction via hierarchically structured learning on relation graphs, *Neural Netw.*, vol. 163, pp. 122–131, 2023.

[180] C. Q. Nguyen, C. Kreatsoulas, and K. M. Branson, Meta-learning GNN initializations for low-resource molecular property prediction, arXiv preprint arXiv: 2003.05996, 2020.

[181] L. Torres, J. P. Arrais, and B. Ribeiro, Few-shot learning via graph embeddings with convolutional networks for low-data molecular property prediction, *Neural Comput. Appl.*, vol. 35, no. 18, pp. 13167–13185, 2023.

[182] H. S. de Ocáriz Borde and F. Barbero, Graph neural network expressivity and meta-learning for molecular property regression, arXiv preprint arXiv: 2209.13410, 2022.

[183] K. P. Ham and L. Sael, Evidential meta-model for molecular property prediction, *Bioinformatics*, vol. 39, no. 10, p. btad604, 2023.

[184] Z. Meng, Y. Li, P. Zhao, Y. Yu, and I. King, Meta-learning with motif-based task augmentation for few-shot molecular property prediction, in *Proc. 2023 SIAM Int. Conf. Data Mining* (*SDM*), Minneapolis-St. Paul Twin Cities, MN, USA, 2023, pp. 811–819.

[185] Z. Guo, C. Zhang, W. Yu, J. Herr, O. Wiest, M. Jiang, and N. V. Chawla, Few-shot graph learning for molecular property prediction, in *Proc. Web Conf. 2021*, Ljubljana, Slovenia, 2021, pp. 2559–2567.

[186] Y. Wang, A. Abuduweili, Q. Yao, and D. Dou, Property-aware relation networks for few-shot molecular property prediction, arXiv preprint arXiv: 2107.07994, 2021.

[187] S. Yao, Z. Feng, J. Song, L. Jia, Z. Zhong, and M.Song, Chemical property relation guided few-shot molecular property prediction, in *Proc. 2022 Int. Joint Conf. Neural Networks* (*IJCNN*), Padua, Italy, 2022, pp. 1–8.

[188] J. Dong, N. N. Wang, Z. J. Yao, L. Zhang, Y. Cheng, D. Ouyang, A. P. Lu, and D. S. Cao, ADMETlab: A platform for systematic ADMET evaluation based on a comprehensively collected ADMET database, *J. Cheminform.*, vol. 10, p. 29, 2018.

[189] D. van Tilborg, A. Alenicheva, and F. Grisoni, Exposing the limitations of molecular machine learning with activity cliffs, *J. Chem. Inf. Model.*, vol. 62, no. 23, pp. 5938–5951, 2022.

[190] Y. Ji, L. Zhang, J. Wu, B. Wu, L. Li, L. K. Huang, T. Xu, Y. Rong, J. Ren, D. Xue, et al., DrugOOD: Out-of-distribution dataset curator and benchmark for AI-aided drug discovery–a focus on affinity prediction problems

with noise annotations, in *Proc. 37th AAAI Conf. Artificial Intelligence*, Washington, DC, USA, 2023, pp. 8023–8031.

[191] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K. R. Müller, Machine learning of accurate energy-conserving molecular force fields, *Sci. Adv.*, vol. 3, no. 5, p. e1603015, 2017.

[192] C. Morris, N. M. Kriege, F. Bause, K. Kersting, P. Mutzel, and M. Neumann, TUDataset: A collection of benchmark datasets for learning with graphs, arXiv preprint arXiv: 2007.08663, 2020.

[193] W. Hu, M. Fey, M. Zitnik, Y. Dong, H. Ren, B. Liu, M. Catasta, and J. Leskovec, Open graph benchmark: Datasets for machine learning on graphs, in *Proc. 34th Int. Conf. Neural Information Processing Systems*, Vancouver, Canada, 2020, pp. 22118–22133.

[194] A. Wojtuch, T. Danel, S. Podlewska, and Ł. Maziarka, Extended study on atomic featurization in graph neural networks for molecular property prediction, *J. Cheminform.*, vol. 15, no. 1, p. 81, 2023.

[195] Z. Zeng, Y. Yao, Z. Liu, and M. Sun, A deep-learning system bridging molecule structure and biomedical text with comprehension comparable to human professionals, *Nat. Commun.*, vol. 13, no. 1, p. 862, 2022.

[196] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, How powerful are graph neural networks? arXiv preprint arXiv: 1810.00826, 2019.

[197] B. Su, D. Du, Z. Yang, Y. Zhou, J. Li, A. Rao, H. Sun, Z. Lu, and J. R. Wen, A molecular multimodal foundation model associating molecule graphs with natural language, arXiv preprint arXiv: 2209.05481, 2022.

[198] X. Tang, A. Tran, J. Tan, and M. B. Gerstein, MolLM: A unified language model to integrate biomedical text with 2D and 3D molecular representations, bioRxiv, doi: 10.1101/2023.11.25.568656.

Taojie Kuang received the BEng degree from South China University of Technology, China in 2022. Currently, he is a PhD candidate at School of Future Technology, South China University of Technology, China, and also with a concurrent affiliation to Peng Cheng Laboratory, China. His research interests include graph neural networks, multi-modal learning, molecular representation, and property prediction.

Pengfei Liu received the BEng degree from Harbin Engineering University, China in 2017 and the MEng degree from Wuhan University, China in 2019. Currently, he is a PhD candidate at School of Computer Science and Engineering, Sun Yat-sen University, China, and also with a concurrent affiliation to Peng Cheng Laboratory, China. His research interests include large language models, multi-modal learning, molecular modeling and design, and visual analytics.

Zhixiang Ren is currently an associate research scientist at Peng Cheng Laboratory, China, and a PhD supervisor at Southern University of Science and Technology, China. He received the PhD degree from University of New Mexico (Albuquerque, USA) in 2018. His research interests include AI for science and multi-modal large AI models, especially “AI+bioinformatics” and AI-aided drug design. He has published over 50 papers with more than 7000 citations. He has been granted 6 national invention patents and 1 software copyright. He has also led the development of one international standard and two industry standards in intelligent computing.