ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Meta Learning with Attention Based FP-GNNs for Few-Shot Molecular Property Prediction

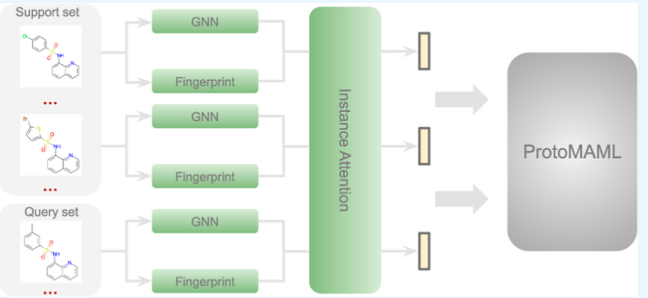
Xiaoliang Qian, Bin Ju, Ping Shen, Keda Yang,* Li Li,* and Qi Liu*

Cite This: *ACS Omega* 2024, 9, 23940−23948

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Molecular property prediction holds significant importance in drug discovery, enabling the identification of biologically active compounds with favorable drug-like properties. However, the low data problem, arising from the scarcity of labeled data in drug discovery, poses a substantial obstacle for accurate predictions. To address this challenge, we introduce a novel architecture, AttFPGNN-MAML, for few-shot molecular property prediction. The proposed approach incorporates a hybrid feature representation to enrich molecular representations and model intermolecular relationships specific to the task. By leveraging ProtoMAML, a meta-learning strategy, our model is trained and adapted to new tasks. Evaluation on two few-shot data sets, MoleculeNet and FS-Mol, demonstrates our method's superior performance in three out of four tasks and across various support set sizes. These results convincingly validate the effectiveness of our method in the realm of few-shot molecular property prediction. The source code is publicly available at https://github.com/sanomics-lab/AttFPGNN-MAML.



INTRODUCTION

One of the primary objectives in the field of drug discovery is to identify biologically active compounds with favorable drug-like properties, including acceptable absorption, distribution, metabolism, excretion, and toxicity (ADME/Tox).[1] Consequently, the development of highly accurate molecular property prediction tools can significantly enhance the efficiency and success rate of drug discovery.

In recent years, the application of deep learning techniques, with a particular emphasis on graph neural networks (GNNs), has led to notable advancements in predicting molecular properties. GNNs are designed to work with molecular graph representations, treating atoms and bonds as nodes and edges. Various deep learning architectures, including graph convolutional networks,[2] graph attention networks,[3] message-passing neural networks (MPNN),[4] AttentiveFP,[5] and directed MPNN,[6] are employed to aggregate node features effectively.

Nonetheless, despite the enormous potential of GNN-based deep learning methods in molecular property prediction, there remains a significant challenge known as the low data problem.[7] This challenge refers to the limited availability of samples for training, which can significantly impact the performance and generalizability of these models. Typically, training a deep learning model for molecular activity/property prediction requires thousands of data points. However, in the context of drug discovery, due to the high cost of experiments and difficulties in data collection, the amount of available data for training is often severely limited. This limitation becomes even more pronounced when dealing with novel drug targets, which often have scarce training data available. This poses challenges for deep learning models in terms of prediction and modeling accuracy because they typically require a large amount of data to achieve optimal performance.

To address the issue of low data, few-shot learning has emerged as a widely used framework. Extensive research efforts have been devoted to exploring few-shot learning methods, with a predominant focus on image data sets,[8,9] and more recently, these methods have been successfully applied to the domain of drug discovery. Few-shot learning methods aim to mitigate the impact of scarce labeled data by leveraging knowledge from a large unlabeled data set or by effectively transferring knowledge from related tasks. The methods within this framework can be categorized into three primary classes:

(1) data augmentation-based methods, which enhance the available samples by generating new and diverse data points through various techniques;[10−12]

(2) embedding-based methods, which learn an embedding space where samples sharing similar properties are positioned closely to each other. For instance, in matching networks,[13] predictions are made based on attention

Received: March 5, 2024
Revised: May 9, 2024
Accepted: May 14, 2024
Published: May 23, 2024

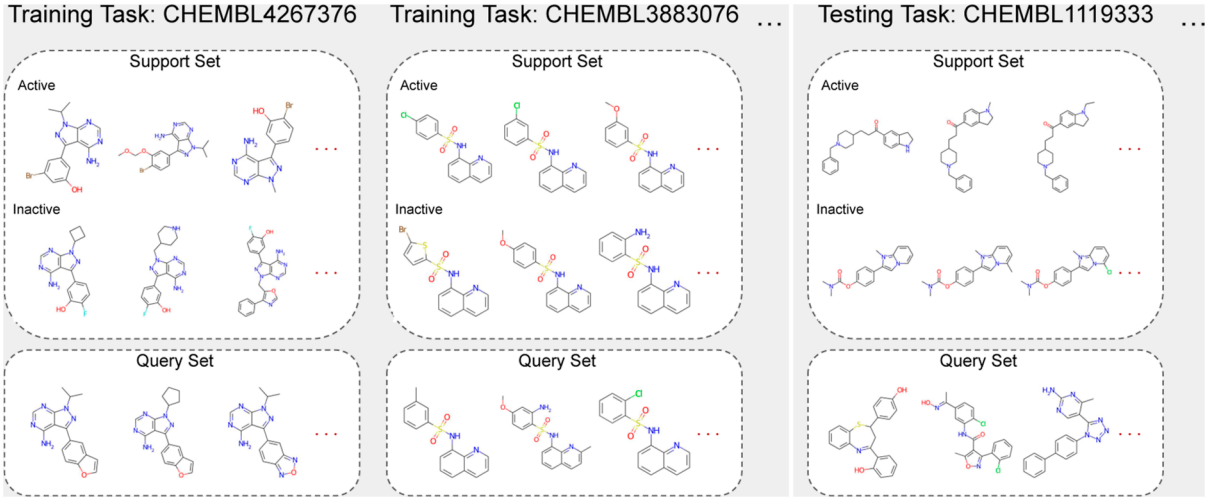


Figure 1. FS-Mol few-shot data set. This is a 2-way K-shot classification problem, where each task corresponds to an assay in ChEMBL.[28] Molecules are labeled as Active or Inactive based on their pIC50 or pEC50 values.

mechanisms operating on the embeddings. Prototype networks[14] generate prototype representations for each class using the embeddings. These methods typically rely on learning similarities between molecules.

(3) Optimization or fine-tuning-based methods, which employ meta-optimizers to efficiently navigate parameter space. For example, model-agnostic meta-learning (MAML)[15] allows meta-optimizers to learn initial weights, which can be adapted to new tasks through a few optimization steps.

Within the field of drug discovery, several few-shot learning methodologies have been proposed. Nguyen et al.[16] conducted an evaluation of the applicability of MAML and its derivatives within GNNs. Guo et al.[17] proposed the *Meta*-MGNN method, which combines MAML with GNNs. This method incorporates self-supervised modules and self-attentive task weighting to enhance few-shot learning performance. Altae-Tran et al.[18] introduced the iterative refinement long short-term memory (IterRefLSTM) method, enabling the mutual sharing of information between the query set and support set, thereby facilitating the iterative update of their embeddings. Property-aware relation networks (PAR)[19] enriched cluster center representations using attention mechanisms and learned relationship graphs between molecules. Chen et al.[20] proposed a framework known as ADKF-IFT, which effectively combines meta-learning and conventional deep kernels. Schimunek et al.[21] introduced the MHNfs method, which leverages large context molecules to enrich molecular representations and achieve superior few-shot molecular property prediction performance. More recently, Ju et al.[22] proposed hierarchically structured learning on relation graphs (HSL-RG), which constructs global relation graphs and utilizes self-supervised learning to acquire transformation-invariant representations of molecules. Meng et al.[23] introduced motif-based task augmentation (MTA) technique to enhance the generalization capability of MAML-based methods, which generate new labeled samples through the retrieval of highly relevant motifs. Stanley et al.[24] recently established a benchmark data set tailored for few-shot drug discovery, and they provided baseline results for a range of methodologies.

Nevertheless, these aforementioned few-shot methods in drug discovery have primarily focused on assessing the effectiveness of various meta-learning algorithms within the field of drug property prediction. There has been a limited emphasis on incorporating of diverse molecular fingerprints to enrich molecular representations and enhance the performance of few-shot molecular property prediction. In this study, we introduce a novel architecture, AttFPGNN-MAML, for few-shot learning in drug discovery, which address this limitation by incorporating a hybrid feature representation that enhances the molecular representation and models intermolecular relationships specific to the given task. Based on this task-specific molecular representation, we use ProtoMAML,[25] a meta-learning strategy, to train our model and adapt to new tasks.

Our approach has been evaluated on two few-shot data sets, MoleculeNet[26] and FS-Mol.[24] In the MoleculeNet benchmark data set, our method outperformed all other approaches on 3 out of 4 tasks. Additionally, our method achieved the best performance on FS-Mol data set at support set sizes of 16, 32, and 64. These results demonstrate the superiority of our method in few-shot learning settings.

PROBLEM SETTING

In few-shot classification, we aim to train models using a diverse set of training tasks and optimize their classification performance across a wide range of testing tasks, including those that have not been previously encountered.[27]

During the few-shot training phase, a multitude of training tasks $D_{\text{train}} = \{T_t\}_{t=1}^{K}$ are presented, with each task T_t comprising two key components: a support set $T_{t,\text{support}}$ and a query set $T_{t,\text{query}}$. The support set is composed of instances equipped with features and corresponding labels, which serve as the basis for model training. Subsequently, the model leverages the features within the query set to predict the labels associated with those instances. During the few-shot testing time, the model encounters an entirely novel and unencountered task T_u. In this context, the model is tasked with predicting labels for the query set $T_{u,\text{query}}$, all the while being granted access to the features and labels of the support set $T_{u,\text{support}}$.

Our primary focus lies in the domains of molecular property classification and molecular activity classification. These domains entail tasks such as forecasting the toxicity and side effects of novel molecules, as well as predicting the bioactivity of new compounds. To illustrate, let us delve into the molecular

https://doi.org/10.1021/acsomega.4c02147
ACS Omega 2024, 9, 23940−23948

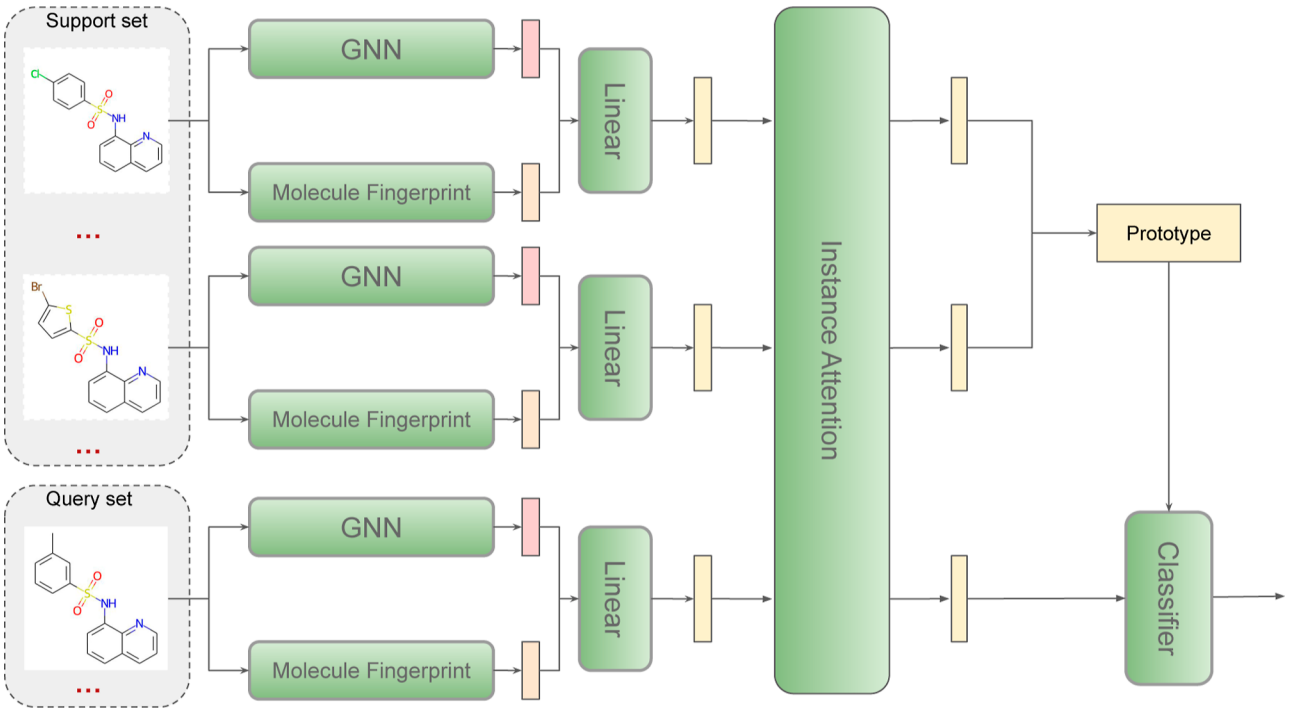


Figure 2. Overall flowchart of our approach. All molecules are fed into a GNN module, a molecule fingerprint module, and a linear layer to obtain fused representations. Then, the instance attention module refines these fused representations within the same task to get task-specific representations. Finally, a classifier layer computes the prediction for the query molecule based on these task-specific representations.

bioactivity prediction task. Here, each task, denoted as $T_i = \{T_{i,\text{support}}, T_{i,\text{query}}\}$, represents an assay. Within both the support and query sets, each sample (x, y) pairs a compound molecule (x) for measurement with a binary experimental label (y) denoting the molecular bioactivity (active or inactive). This task can be categorized as a 2-way K-shot problem, where the "2-way" signifies that each task involves two classes, and "*K*-shot" denotes that we sample *K* molecules for each of these classes to compose the support set. Figure 1 illustrates a typical few-shot learning scenario in the FS-Mol data set.

METHODOLOGY

The flowchart of our approach is presented in Figure 2. Within this approach, the molecules in the support set and query set initially undergo the GNN module and the molecule fingerprint module to obtain two different molecular feature representations. Subsequently, these two molecular representations are concatenated and fed into a fully connected layer to produce a fused molecular representation. Following this, the representations of all molecules within the same task are further refined through the instance attention module, yielding task-specific representations of the molecules. Finally, based on the obtained task-specific molecular representations, the entire model is trained using the ProtoMAML meta-learning strategy. In the subsequent part of this section, we will provide a comprehensive introduction to each module and training strategy employed within the model.

Graph Neural Networks. A molecule can be described as an undirected graph $G = (V, E)$, where V represents the nodes (atoms), and the number of nodes is denoted as $|V| = N$. E represents the set of edges (bonds), and the number of edges is $|E| = M$. Each node $v_i \in V$ and edge $e_{i,j} = (v_i, v_j) \in E$ is initially assigned with attributes $x_i \in \mathbb{R}^{dn}$ and $e_{i,j} \in \mathbb{R}^{de}$, where dn and de represent the feature dimensions corresponding to nodes and edges.

In the context of predicting molecular properties, the majority of GNN-based models adhere to a message-passing paradigm. This paradigm relies on three key functions—message passing function, aggregation function, and update function—to iteratively extract atomic features. The expression for the k-th layer of the message-passing paradigm is as follows

$$m_i^k = \text{aggregate}^k(\{\text{message}^k(h_i^{k-1}, h_j^{k-1}, e_{ji}): j \in N(i)\})$$

$$h_i^k = \text{update}^k(h_i^{k-1}, m_i^k)$$

Here, message, aggregate, and update represent the message-passing function, aggregation function, and update function, respectively. h_i^k represents the hidden state of node i within the k-th layer. e_{ji} represents the feature vector associated with the edge connecting node j to node i. The set $N(i)$ refers to the collection of neighboring nodes linked to node i. Additionally, employing a readout function enables the acquisition of the representation of the complete graph

$$h^G = \text{readout}(\{h_i^K: v_i \in G\})$$

where K represents the total number of iterations and readout denotes a perturbation-invariant function over a set of nodes.[29] Ultimately, a global embedding representation of the molecule is achieved via GNNs.

Molecular Fingerprint. To complement the potentially missing chemical and structural information in graph-based representations and provide a more comprehensive depiction, we introduce an additional molecular fingerprint module. Molecular fingerprints serve as a method for encoding the structures of compounds, effectively capturing the diverse and subtle structural features inherent in molecules.

https://doi.org/10.1021/acsomega.4c02147
ACS Omega 2024, 9, 23940−23948

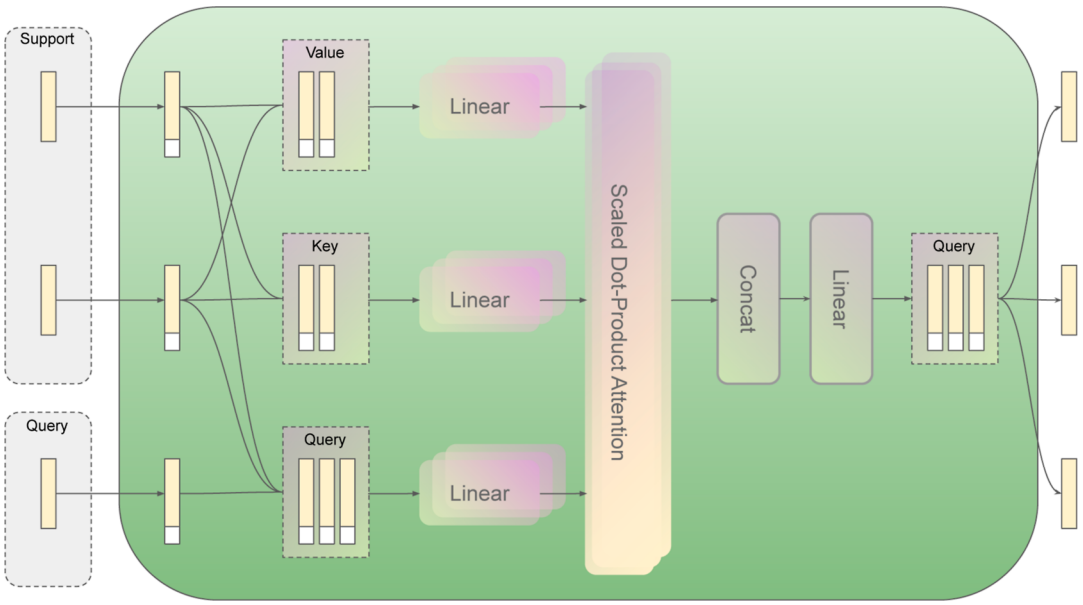


Figure 3. Instance attention module. This module captures the relationships between different molecules within the same task, resulting in task-specific molecular representations.

We use the mixed fingerprints as described in the FP-GNN.[30−32] Our model employs three distinct fingerprints [MACCS fingerprint,[33] Pharmacophore extended reduced graph (ErG) fingerprint,[34] and PubChem fingerprint[35] due to their complementary and comprehensive representation of molecular features.[36] The following are brief descriptions of these three fingerprints.

(1) MACCS fingerprint: It is a substructure-based molecular fingerprint. MACCS fingerprint includes many predefined SMARTS patterns, which is meaningful for drug discovery. In this study, we have selected a short variant with 1 + 166 bits.
(2) Pharmacophore ErG fingerprint: it is a two-dimensional pharmacophore fingerprint that encodes molecular properties using the ErG method.
(3) PubChem fingerprint: It is an 881-bit substructure-based fingerprint that extensively covers a diverse array of substructures.

We combine these three types of fingerprints together to create a comprehensive molecular fingerprint representation

$$\mathrm{FP} = \mathrm{concat}(\mathrm{FP}_{\mathrm{MACCS}}, \mathrm{FP}_{\mathrm{PubChem}}, \mathrm{FP}_{\mathrm{ErG}})$$

After obtaining the GNN molecular graph representation and molecular fingerprint representation mentioned above, we use a multilayer feedforward network to obtain the fused molecular representation

$$P = MLP(\mathrm{concat}(h^{\mathrm{G}}, \mathrm{FP}))$$

Instance Attention. The Instance Attention module is illustrated in Figure 3. In this module, we use multihead attention[37] on the fused molecular representations to capture the relationships between molecules within the same task, resulting in task-specific molecular representations.

Multihead attention is an attention mechanism module that runs the attention mechanism multiple times in parallel. The independently computed attention outputs are then concatenated and linearly transformed to the desired dimension. Intuitively, multihead attention allows us to focus on different aspects of the sequence in various ways, such as capturing long-term and short-term dependencies.

$$\mathrm{MultiHead}(Q, K, V) = [\mathrm{head}_1; \ldots; \mathrm{head}_h]W^O$$

$$\mathrm{head}_i = \mathrm{attention}(QW^Q, KW^K, VW^V)$$

The attention employed here is the scaled dot-product attention. In this attention mechanism, the output is derived as a weighted sum of the values, with the weight assigned to each value determined by the dot product of the query with the keys

$$\mathrm{attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{n}}\right)V$$

In the above formula, all the W (W^O, W^Q, W^K, W^V) are learnable parameters. In our work, we set K and V as the fused features of the molecules in the support set, i.e., P_{support} and set Q as the fused features of all molecules in the query set and support set, i.e., contact (P_{query}, P_{support}).

Additionally, to consider the influence of molecular categories, we add label embeddings to each molecular representation as MHNfs.[21] The molecules in the support set are concatenated with all 1 or all −1 vectors according to their respective categories, while the query molecules are concatenated with a vector of all 0 s.

ProtoMAML. After obtaining the task-specific molecular representation $g(x, \theta)$, we use a linear classification layer to classify the query samples.

$$p(y|x, S) = \mathrm{softmax}\,(b + Wg(x, \theta))$$

The entire model is trained using the ProtoMAML meta-learning algorithm.

During the training phase, the learning algorithm first compute new parameters using the support set of a particular task. Then, it uses these new parameters to calculate the loss on the query set for that task, which is used to update the model parameters. This approach effectively learns an initial model parametrization that can quickly adapt to new tasks.

https://doi.org/10.1021/acsomega.4c02147
ACS Omega 2024, 9, 23940−23948

During the inference phase, the model parameters can be updated once or a few times based on the support set of a new task. Subsequently, these updated parameters are then used for making predictions on the query set.

In our approach, we initialize the classification layer parameters based on the logits calculated in ProtoNet. In ProtoNet, the classification probability is calculated as follows

$$p(y = c|x) = \mathrm{softmax}(-d_\phi(g(x, \theta), v_c)) = \frac{\exp - d_\phi(g(x, \theta), v_c)}{\sum_{c' \in C} -d_\phi(g(x, \theta), v_{c'})}$$

In the above formula, d represents the distance metric function. When using the Mahalanobis distance as the distance metric, the distance between query representation and prototypical representation v_c can be written as

$$d_\phi(g(x, \theta), v_c)) = (g(x, \theta) - v_c)^T \Sigma_c^{-1}(g(x, \theta) - v_c) = (g(x, \theta)\Sigma_c^{-1}(g(x, \theta) - 2v_c^T \Sigma_c^{-1} g(x, \theta) + v_c^T \Sigma_c^{-1} v_c \quad (1)$$

where Σ_c refers to the covariance matrix of molecular features belonging to class c. So, here we initialize the parameter W in the classification layer as $-2v_c^T\Sigma^{-1}$ and initialize parameter b as $v_c^T\Sigma^{-1}v_c$. Note that we ignore the first term on the right-hand side of the Mahalanobis distance expansion.

RESULTS AND DISCUSSION

Benchmarking on MoleculeNet. *Benchmark and Baselines.* We conduct experiments on four commonly used public benchmark data sets for few-shot molecular property prediction (Tox21, Sider, MUV, Toxcast) to assess the performance of various models. These four data sets are part of MoleculeNet,[26] a comprehensive benchmark designed for evaluating machine learning approaches in the domain of molecular-related tasks.

Tox21—Designed to predict the Toxicity of Molecules. It is a publicly accessible database established by the "Toxicology in the 21st Century" program, aiming at assessing the toxicity of various compounds. This data set comprises qualitative toxicity measurements on 12 distinct biological targets, including nuclear receptors and stress response pathways. Initially, Tox21 was utilized in IterRefLSTM[18] for the evaluation of few-shot activity prediction, where the authors employed the first 9 assays for meta-training and the last 3 assays for meta-testing.

SIDER—Used to predict potential Side Effects of Drugs. It is a database that compiles marketed drugs and adverse drug reactions, categorized into 27 system organ classes. Similar to Tox21, SIDER was also initially employed in IterRefLSTM[18] for evaluating few-shot activity prediction. In IterRefLSTM,[18] the researchers utilized the first 21 assays from the original data set for meta-training and reserved the last 6 assays for meta-testing.

MUV—Used for the prediction of Molecular Activities. It is a subset derived from PubChem BioAssay through refined nearest neighbor analysis, and it is specifically crafted for validating virtual screening techniques. This data set was also initially used in IterRefLSTM[18] for evaluating few-shot activity prediction. In their evaluation, they selected the first 12 assays from the original data set for meta-training and designated the final 5 assays for meta-testing.

ToxCast—Designed to predict the Toxicological Effects of Compounds. It contains toxicology data for an extensive library of compounds obtained through in vitro high-throughput screening, involving experiments across more than 600 tasks. The data set was initially employed for few-shot activity prediction evaluation in PAR,[19] where the authors randomly split the data set into 450 training assays and 167 testing assays.

We use the same task splits as describe in previous works[18−20] for our analysis and present the statistics of the four few-shot benchmarks in Table 1.

Table 1. Statistics of Four MoleculeNet Few-Shot Data sets

	# compounds	# tasks	# training tasks	# test tasks
Tox21	8014	12	9	3
SIDER	1427	27	21	6
MUV	93127	17	12	5
ToxCast	8615	617	450	167

We conduct a comparative analysis of our approach against two categories of baselines across the four few-shot MoleculeNet benchmark tasks.

(1) Methods without a pretrained molecular graph encoder, such as HSL-RG$^-$,[22] ADKF-IFT,[20] PAR,[19] IterRefLSTM,[18] EGNN,[38] TPN,[39] MAML,[15] ProtoNet,[14] and Siamese.[40]

(2) Methods with a pretrained molecular graph encoder, such as Pre-PAR + +MTA,[23] HSL-RG,[22] Pre-GNN,[41] Meta-MGNN,[17] Pre-PAR,[19] and Pre-ADKF-IFT.[20] It is worth noting that all methods within this category employ a pretrained GIN, and the pretrained GIN weights are provided by Hu et al.[41]

Evaluation Procedure and Performance. Following Chen et al.,[20] we use AUROC (Area Under the ROC curve) as the task level metric. We present the average performance of each compared method over ten runs with different random seeds. In all these experiments, the support set size is set to 20 (i.e., 2-way 10-shot). We do not conduct one-shot learning as it is not feasible for practical drug discovery tasks. All baseline results are sourced from the work of Chen et al.[20]

Table 2 displays the performance of AttFPGNN-MAML and other baseline models on the MoleculeNet few-shot data sets. The results demonstrate that our method significantly outperforms the previous approaches in terms of the AUROC metric when not utilizing a pretrained molecular graph encoder, with the exception of the MUV data set. On Tox21, SIDER, and ToxCast, our method achieves remarkable improvements of 1.88, 9.76, and 4.97%, respectively, compared to the previous state-of-the-art (SOTA) methods. This enhancement underscores the efficacy of our proposed methodology. On the MUV data set, our method outperforms all other methods except ADKF-IFT. ADKF-IFT exhibits exceptionally superior performance on the MUV data set, surpassing all other methods, including ours. As mentioned in IterRefLSTM,[18] the positive samples in the MUV data set are structurally distinct and dissimilar to each other, which poses a challenge for certain few-shot methods to effectively leverage the structural similarities between molecules for predicting the activity of new molecules. Conversely, ADKF-IFT, as an extension of the basic machine learning method Deep Kernel Learning, holds a distinct

https://doi.org/10.1021/acsomega.4c02147
ACS Omega 2024, 9, 23940−23948

Table 2. Mean AUROC and Standard Deviation of Various Methods Across Four MoleculeNet Few-Shot Benchmark Tasks

	Tox21 (8014)	SIDER (1,427)	MUV (93,127)	ToxCast (8,615)
Siamese[a]	80.40 ± 0.35	71.10 ± 4.32	59.59 ± 5.13	
ProtoNet[a]	74.98 ± 0.32	64.54 ± 0.89	65.88 ± 4.11	63.70 ± 1.26
MAML[a]	80.21 ± 0.24	70.43 ± 0.76	63.90 ± 2.28	66.79 ± 0.85
TPN[a]	76.05 ± 0.24	67.84 ± 0.95	65.22 ± 5.82	62.74 ± 1.45
EGNN[a]	81.21 ± 0.16	72.87 ± 0.73	65.20 ± 2.08	63.65 ± 1.57
IterRefLSTM[a]	81.10 ± 0.17	69.63 ± 0.31	45.56 ± 5.12	
PAR[a]	82.06 ± 0.12	74.68 ± 0.31	66.48 ± 2.12	69.72 ± 1.63
ADKF-IFT[a]	82.43 ± 0.60	67.72 ± 1.21	**98.18 ± 3.05**	72.07 ± 0.81
HSL-RG−[b]	80.95 ± 0.26	74.66 ± 0.52	70.38 ± 1.35	70.70 ± 1.02
AttFPGNN-MAML	**84.31 ± 0.22**	**84.44 ± 0.08**	79.67 ± 0.91	**77.04 ± 0.15**
pre-GNN[a]	82.14 ± 0.08	73.96 ± 0.08	67.14 ± 1.58	73.68 ± 0.74
meta-MGNN[a]	82.97 ± 0.10	75.43 ± 0.21	68.99 ± 1.84	
pre-PAR[a]	84.93 ± 0.11	78.08 ± 0.16	69.96 ± 1.37	75.12 ± 0.84
pre-ADKF-IFT[a]	86.06 ± 0.35	70.95 ± 0.60	**95.74 ± 0.37**	76.22 ± 0.13
HSL-RG[b]	85.56 ± 0.28	78.99 ± 0.33	71.26 ± 1.08	76.00 ± 0.81
pre-PAR + MTA[c]	**86.69 ± 0.73**	79.73 ± 0.88	71.49 ± 1.06	76.27 ± 1.12
pre-AttFPGNN-MAML	86.12 ± 0.26	**84.68 ± 0.01**	80.21 ± 0.29	**78.15 ± 0.06**

[a]Results from Chen et al.[20] [b]Results from Ju et al.[22] [c]Results from Meng et al.[23]

advantage in this scenario. Additionally, when incorporating a pretrained molecular graph encoder, the performance of our method is further improved, and it still maintains a certain advantage over other methods that employ the same pretrained encoder in terms of the AUROC metric.

Benchmarking on FS-Mol. *Benchmark and Baselines.* The FS-Mol data set, extracted from ChEMBL27, is a rich and diverse collection of 489,133 measurements, representing 233,786 unique chemical compounds and spanning 5120 distinct tasks. With an average of 94 data points per task, the data set strikes an excellent balance between activity and inactivity, with an average ratio of active molecules to inactive molecules close to 1. To ensure effective training, validation, and testing, the FS-Mol benchmark data set is carefully divided into 4938 training tasks, 40 validation tasks, and 157 test tasks. The data set further includes precomputed features such as extended connectivity fingerprints (ECFP)[42] and key molecular physical descriptors defined by RDKit,[43] providing an efficient and comprehensive basis for molecular property prediction and drug discovery research.

We compared our method with a range of baseline methods, including kNN, Random Forest,[24,44] GNN-ST,[4,24] GNN-MT,[24,45] MAT,[24,46] GNN-MAML,[15,24] ProtoNet,[14,24] PAR,[19] and ADKF-IFT.[20]

Evaluation Procedure and Performance. We adopt the evaluation protocol established by Stanley et al.[24] The task-level metric is ΔAUPRC, which represents the difference in area under the precision−recall curve between the classifier and a random classifier. This evaluation metric emphasizes more on the enhancement achieved by the learned classifier in comparison to a trivial classifier.

$$\Delta\text{AUPRC} = \text{AUPRC}(f(T_{u,\text{query}})) - \frac{\#\text{active compounds in } T_{u,\text{query}}}{|T_{u,\text{query}}|}$$

We report the average performance of various methods across support set sizes of 16, 32, 64, 128, and 256. The results, as shown in Figure 4, clearly demonstrate that our approach outperforms Random Forest, GNN-ST, GNN-MT, MAT, GNN-MAML, ProtoNet, and PAR methods by a significant

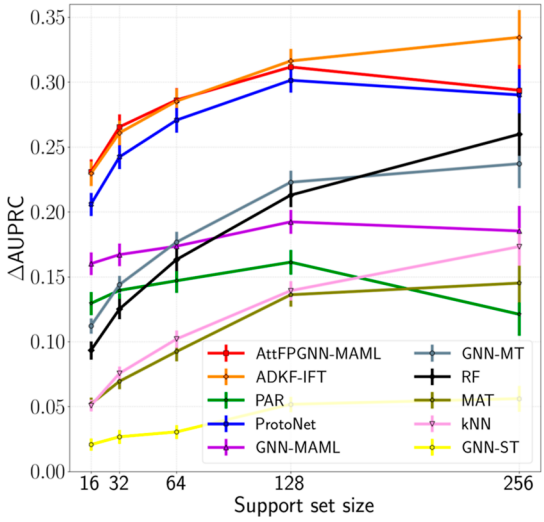


Figure 4. Line chart of the performance of different methods across 157 tasks in the FS-Mol data set. Please refer to Figures S1−S5 in Supporting Information for more data.

margin. Notably, when support set sizes are relatively small (16, 32, 64), our method performs on par with the SOTA approach ADKF-IFT. However, as the support set size increases, our method's performance tends to lag behind ADKF-IFT, particularly when the support set size reaches 256. At a support set size of 256, some tasks within the testing set cannot be used for meta-testing due to inadequate data points. Ultimately, out of the 157 testing tasks, only 43 tasks have a sufficient number of sample points available for meta-testing at this point. This might be the reason for the decline in the performance of our method at this specific support set size. To complement these observations, we present box plots illustrating the performance of various methods in Supporting Information Figures S1−S5. Besides, we also report the mean rank in comparison with other methods, which is shown in Table 3. The mean rank is calculated by autorank,[47] following Demšar.[48]

Additionally, to account for the diversity of the tasks under consideration, we conduct a more extensive examination of the classification performance variations across different enzyme

https://doi.org/10.1021/acsomega.4c02147
ACS Omega 2024, 9, 23940−23948

Table 3. Mean Ranks of All Compared Methods in Terms of Their Performance on All FS-Mol Test Tasks

	support set size				
method	16	32	64	128	256
GNN-ST	8.66	8.85	9.05	9.19	9.44
kNN	8.16	7.86	7.73	7.70	7.02
MAT	7.89	7.95	7.79	7.53	7.40
random forest	6.08	5.91	5.47	4.98	3.58
GNN-MT	5.70	5.70	5.65	5.23	5.05
GNN-MAML	4.99	5.49	5.90	6.28	6.92
ProtoNet	3.26	2.97	2.81	2.69	3.10
PAR	5.98	6.38	6.59	7.03	8.38
ADKF-IFT	2.16	2.13	2.14	**2.11**	**1.27**
AttFPGNN-MAML	**2.12**	**1.76**	**1.86**	2.26	2.85

commission numbers (EC numbers) within the FS-Mol test tasks. The EC number is a numerical classification scheme for enzymes, delineating their categorization based on the specific chemical reactions they catalyze. We categorize the evaluation results into distinct subclasses based on different EC numbers. The categorized results are presented in Table 4. It reveals that our method consistently outperforms the previous baseline models across different EC categories, achieving the SOTA performance at small support set sizes (16, 32, 64).

Ablation Study. We conduct ablation experiments on 10-shot tasks from SIDER. This study involves a comparative analysis between the original model and variants from which the Molecular Fingerprint module or Instance Attention module is excluded: (i) w/o InsAtt: w/o applying the Instance Attention module; (ii) w/o MolFP: w/o applying the Molecular Fingerprint module; and (iii) w/o InsAtt + MolFP: w/o applying Instance Attention module and Molecular Fingerprint module. As depicted in Figure 5, the removal of Molecular Fingerprint module and Instance Attention module results in a discernible degree of performance degradation. This substantiates the assertion that the introduction of the Molecular Fingerprint module and Instance Attention module has a positive impact on the overall efficacy of the model.

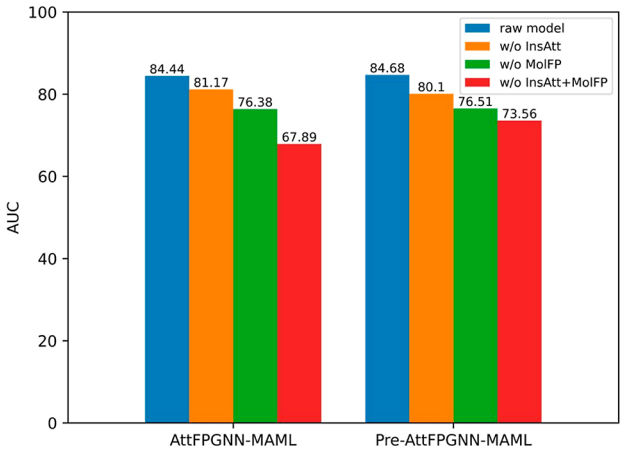


Figure 5. Ablation study on 10-shot tasks from SIDER. The model's performance exhibits distinct levels of deterioration upon the removal of the molecular fingerprint and instance attention modules, thereby substantiating the effectiveness of these two components.

CONCLUSIONS

In this study, we propose a novel approach for few-shot molecular property prediction. This method combines the GNN with mixed fingerprints to generate more comprehensive molecular representations. Additionally, Instance Attention is employed to obtain task-specific representations for molecules across different tasks. Our approach has been tested on two distinct few-shot benchmark data sets, MoleculeNet and FS-Mol, demonstrating excellent performance. Furthermore, we validate the effectiveness of the Molecular Fingerprint and Instance Attention module through ablation experiments conducted on the MoleculeNet data set.

We do not provide interpretability of our model, which can be problematic when trying to identify potential drug candidates or understand molecular interactions. In the future, we will try to demonstrate the interpretability of our model and verify the effectiveness of the model in real projects. We plan to compute the attention scores within the InstanceAttention module of the meta-learned model. This involves assessing whether the attention scores, which reflect the relationships between query molecules and support set molecules, are consistent with expert evaluations in medicinal chemistry. Also, we intend to explore the existing GNN interpretability tools to identify crucial substructures within the molecules that contribute significantly to classification.

ASSOCIATED CONTENT

Data Availability Statement

The data sets used in this study and the source code for AttFPGNN-MAML are publicly available at https://github.com/sanomics-lab/AttFPGNN-MAML.

Table 4. Performance of Different Methods with a Support Set Size of 16[a]

method	all [157]	kin. [125]	hydrol. [20]	oxid.[7]
GNN-ST	0.021 ± 0.005	0.013 ± 0.004	0.062 ± 0.019	0.013 ± 0.019
kNN	0.051 ± 0.005	0.046 ± 0.005	0.085 ± 0.019	0.043 ± 0.018
MAT	0.052 ± 0.005	0.043 ± 0.005	0.095 ± 0.019	0.063 ± 0.024
random forest	0.093 ± 0.007	0.082 ± 0.007	0.158 ± 0.028	0.081 ± 0.032
GNN-MT	0.112 ± 0.006	0.113 ± 0.006	0.129 ± 0.025	0.046 ± 0.013
GNN-MAML	0.160 ± 0.009	0.178 ± 0.009	0.106 ± 0.024	0.046 ± 0.023
ProtoNet	0.206 ± 0.009	0.217 ± 0.009	0.196 ± 0.031	0.086 ± 0.029
PAR	0.129 ± 0.009	0.147 ± 0.010	0.068 ± 0.021	0.008 ± 0.005
ADKF-IFT	**0.230 ± 0.009**	**0.243 ± 0.010**	0.213 ± 0.029	0.103 ± 0.036
AttFPGNN-MAML	**0.231 ± 0.010**	**0.243 ± 0.010**	**0.215 ± 0.031**	**0.111 ± 0.032**

[a]Results are broken down by the EC category. Please refer to Table S3−S7 in Supporting Information for more detailed data.

https://doi.org/10.1021/acsomega.4c02147
ACS Omega 2024, 9, 23940−23948

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.4c02147.

Model hyperparameters; performance of different methods within FS-Mol sub EC categories across various support set sizes; and box plots of the performance of different methods within the FS-Mol data set across various support set sizes (PDF)

AUTHOR INFORMATION

Corresponding Authors

Keda Yang − *Shulan International Medical College, Zhejiang Shuren University, Hangzhou 310015, China;* orcid.org/0009-0001-9368-9737; Email: kdyang@zjsru.edu.cn

Li Li − *Department of Hepatobiliary Surgery, The First People's Hospital of Kunming, Kunming 650034, China;* Email: lilikm26@163.com

Qi Liu − *Key Laboratory of Spine and Spinal Cord Injury Repair and Regeneration (Tongji University), Ministry of Education, Orthopaedic Department of Tongji Hospital, Frontier Science Center for Stem Cell Research, Bioinformatics Department, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China; Translational Medical Center for Stem Cell Therapy and Institute for Regenerative Medicine, Shanghai East Hospital, Frontier Science Center for Stem Cell Research, Bioinformatics Department, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China; Shanghai Research Institute for Intelligent Autonomous Systems, Shanghai 201804, China;* Email: qiliu@tongji.edu.cn

Authors

Xiaoliang Qian − *Translational Medical Center for Stem Cell Therapy and Institute for Regenerative Medicine, Shanghai East Hospital, Frontier Science Center for Stem Cell Research, Bioinformatics Department, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China; SanOmics AI Co., Ltd., Hangzhou 311103, China;* orcid.org/0009-0002-8363-1264

Bin Ju − *SanOmics AI Co., Ltd., Hangzhou 311103, China; State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310009, China*

Ping Shen − *State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310009, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c02147

Funding

This work was supported by the National Key Research and Development Program of China (Grant No. 2021YFF1201200, No. 2021YFF1200900), National Natural Science Foundation of China (Grant No. 32341008, 62088101), Shanghai Pilot Program for Basic Research, Shanghai Science and Technology Innovation Action Plan-Key Specialization in Computational Biology, Shanghai Shuguang Scholars Project, Shanghai Excellent Academic Leader Project, Shanghai Municipal Science and Technology Major Project (Grant No. 2021SHZDZX0100) and Fundamental Research Funds for the Central Universities.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank the colleagues of SanOmics AI Co., Ltd. for their invaluable assistance during the course of this research. The authors also thank the members of the Bioinformatics Department of Tongji University for their helpful advice and guidance.

REFERENCES

(1) Lv, Q.; Chen, G.; Yang, Z.; Zhong, W.; Chen, C. Y.-C. Meta learning with graph attention networks for low-data drug discovery. *IEEE Transact. Neural Networks Learn. Syst.* **2024**, 1−13.

(2) Kipf, T. N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*; Cornell University, 2016.

(3) Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. In *International Conference on Learning Representations*; ICLR, 2018.

(4) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural message passing for quantum chemistry. In *International Conference on Machine Learning*; PMLR, 2017; pp 1263−1272.

(5) Xiong, Z.; Wang, D.; Liu, X.; Zhong, F.; Wan, X.; Li, X.; Li, Z.; Luo, X.; Chen, K.; Jiang, H.; Zheng, M. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *J. Med. Chem.* **2020**, *63*, 8749−8760.

(6) Yang, K.; Swanson, K.; Jin, W.; Coley, C.; Eiden, P.; Gao, H.; Guzman-Perez, A.; Hopper, T.; Kelley, B.; Mathea, M.; Palmer, A.; Settels, V.; Jaakkola, T.; Jensen, K.; Barzilay, R. Analyzing learned molecular representations for property prediction. *J. Chem. Inf. Model.* **2019**, *59*, 3370−3388.

(7) Silva-Mendonça, S.; Vitória, A. R. d. S.; Lima, T. W. d.; Galvão-Filho, A. R.; Andrade, C. H. Exploring new horizons: Empowering computer-assisted drug design with few-shot learning. *Artif. Intell. Life Sci.* **2023**, *4*, 100086.

(8) Bendre, N.; Marín, H. T.; Najafirad, P. Learning from few samples: A survey. 2020, arXiv:2007.15484. arXiv preprint. https://doi.org/10.48550/arXiv.2007.15484.

(9) Wang, Y.; Yao, Q.; Kwok, J. T.; Ni, L. M. Generalizing from a few examples: A survey on few-shot learning. *ACM Comput. Surv.* **2021**, *53*, 1−34.

(10) Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning*; PMLR, 2020; pp 1597−1607.

(11) Zhao, A.; Balakrishnan, G.; Durand, F.; Guttag, J. V.; Dalca, A. V. Data augmentation using learned transformations for one-shot medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; CVPR, 2019; pp 8543−8553.

(12) Antoniou, A.; Storkey, A. Assume, augment and learn: Unsupervised few-shot meta-learning via random labels and data augmentation. 2019, arXiv:1902.09884. arXiv preprint. https://doi.org/10.48550/arXiv.1902.09884.

(13) Vinyals, O.; Blundell, C.; Lillicrap, T.; kavukcuoglu, k.; Wierstra, D. Matching networks for one shot learning. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 3630−3638.

(14) Snell, J.; Swersky, K.; Zemel, R. Prototypical networks for few-shot learning. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4077−4087.

(15) Finn, C.; Abbeel, P.; Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*; PMLR, 2017; pp 1126−1135.

(16) Nguyen, C. Q.; Kreatsoulas, C.; Branson, K. M. Meta-Learning GNN Initializations for Low-Resource Molecular Property Prediction. In *4th Lifelong Machine Learning Workshop at ICML 2020*; Cornell University, 2020.
(17) Guo, Z.; Zhang, C.; Yu, W.; Herr, J.; Wiest, O.; Jiang, M.; Chawla, N. V. Few-shot graph learning for molecular property prediction. In *Proceedings of the Web Conference 2021*; ACM, 2021; pp 2559−2567.
(18) Altae-Tran, H.; Ramsundar, B.; Pappu, A. S.; Pande, V. Low data drug discovery with one-shot learning. *ACS Cent. Sci.* **2017**, *3*, 283−293.
(19) Wang, Y.; Abuduweili, A.; Yao, Q.; Dou, D. Property-aware relation networks for few-shot molecular property prediction. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 17441−17454.
(20) Chen, W.; Tripp, A.; Lobato, J. M. H. Meta-learning Adaptive Deep Kernel Gaussian Processes for Molecular Property Prediction. 2023, arXiv:2205.02708. arXiv preprint. https://doi.org/10.48550/arXiv.2205.02708.
(21) Schimunek, J.; Seidl, P.; Friedrich, L.; Kuhn, D.; Rippmann, F.; Hochreiter, S.; Klambauer, G. Context-enriched molecule representations improve few-shot drug discovery. 2023, arXiv:2305.09481. arXiv preprint. https://doi.org/10.48550/arXiv.2305.09481.
(22) Ju, W.; Liu, Z.; Qin, Y.; Feng, B.; Wang, C.; Guo, Z.; Luo, X.; Zhang, M. Few-shot molecular property prediction via Hierarchically Structured Learning on Relation Graphs. *Neural Network.* **2023**, *163*, 122−131.
(23) Meng, Z.; Li, Y.; Zhao, P.; Yu, Y.; King, I. Meta-Learning with Motif-based Task Augmentation for Few-Shot Molecular Property Prediction. In *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*; SIAM, 2023; pp 811−819.
(24) Stanley, M.; Bronskill, J. F.; Maziarz, K.; Misztela, H.; Lanini, J.; Segler, M.; Schneider, N.; Brockschmidt, M. Fs-mol: A few-shot learning dataset of molecules. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*; NeurIPS, 2021.
(25) Triantafillou, E.; Zhu, T.; Dumoulin, V.; Lamblin, P.; Evci, U.; Xu, K.; Goroshin, R.; Gelada, C.; Swersky, K.; Manzagol, P.-A.; Larochelle, H. Meta-Dataset: A Dataset of Datasets for Learning to Learn from Few Examples. *International Conference on Learning Representations*; Cornell University, 2019.
(26) Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; Pande, V. MoleculeNet: a benchmark for molecular machine learning. *Chem. Sci.* **2018**, *9*, 513−530.
(27) Vella, D.; Ebejer, J.-P. Few-shot learning for low-data drug discovery. *J. Chem. Inf. Model.* **2023**, *63*, 27−42.
(28) Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M.; Mosquera, J.; Mutowo, P.; Nowotka, M.; et al. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* **2019**, *47*, D930−D940.
(29) Xu, K.; Hu, W.; Leskovec, J.; Jegelka, S. How powerful are graph neural networks? 2018, arXiv:1810.00826. arXiv preprint. https://doi.org/10.48550/arXiv.1810.00826.
(30) Cai, H.; Zhang, H.; Zhao, D.; Wu, J.; Wang, L. FP-GNN: a versatile deep learning architecture for enhanced molecular property prediction. *Briefings in bioinformatics* **2022**, *23*, bbac408.
(31) Ai, D.; Wu, J.; Cai, H.; Zhao, D.; Chen, Y.; Wei, J.; Xu, J.; Zhang, J.; Wang, L. A multi-task FP-GNN framework enables accurate prediction of selective PARP inhibitors. *Front. Pharmacol* **2022**, *13*, 971369.
(32) Zhang, H.; Huang, J.; Chen, R.; Cai, H.; Chen, Y.; He, S.; Xu, J.; Zhang, J.; Wang, L. Ligand- and structure-based identification of novel CDK9 inhibitors for the potential treatment of leukemia. *Bioorg. Med. Chem.* **2022**, *72*, 116994.
(33) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273−1280.
(34) Stiefl, N.; Watson, I. A.; Baumann, K.; Zaliani, A. ErG: 2D pharmacophore descriptions for scaffold hopping. *J. Chem. Inf. Model.* **2006**, *46*, 208−220.
(35) Bolton, E. E.; Wang, Y.; Thiessen, P. A.; Bryant, S. H. *PubChem: Integrated Platform of Small Molecules and Biological Activities*; Elsevier, 2008; Vol. *4*, pp 217−241.
(36) Shen, W. X.; Zeng, X.; Zhu, F.; Wang, Y. L.; Qin, C.; Tan, Y.; Jiang, Y. Y.; Chen, Y. Z. Out-of-the-box deep learning prediction of pharmaceutical properties by broadly learned knowledge-based molecular representations. *Nat. Mach. Intell.* **2021**, *3*, 334−343.
(37) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998−6008.
(38) Kim, J.; Kim, T.; Kim, S.; Yoo, C. D. Edge-labeling graph neural network for few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; CVPR, 2019; pp 11−20.
(39) Liu, Y.; Lee, J.; Park, M.; Kim, S.; Yang, E.; Hwang, S. J.; Yang, Y. Learning to propagate labels: Transductive propagation network for few-shot learning. In *7th International Conference on Learning Representations, ICLR 2019*; Cornell University, 2019.
(40) Koch, G.; Zemel, R.; Salakhutdinov, R. Siamese neural networks for one-shot image recognition. *ICML Deep Learning Workshop*; Lille, 2015.
(41) Hu, W.; Liu, B.; Gomes, J.; Zitnik, M.; Liang, P.; Pande, V.; Leskovec, J. Strategies For Pre-training Graph Neural Networks. In *International Conference on Learning Representations*; ICLR, 2020.
(42) Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742−754.
(43) Greg, L. *Rdkit: A Software Suite for Cheminformatics. Computational Chemistry, and Predictive Modeling*, 2013.
(44) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5−32.
(45) Alain, G.; Bengio, Y. Understanding intermediate layers using linear classifier probes. 2016, arXiv:1610.01644. arXiv preprint. https://doi.org/10.48550/arXiv.1610.01644.
(46) Maziarka, Ł.; Danel, T.; Mucha, S.; Rataj, K.; Tabor, J.; Jastrzębski, S. Molecule attention transformer. 2020, arXiv:2002.08264. arXiv preprint. https://doi.org/10.48550/arXiv.2002.08264.
(47) Herbold, S. Autorank: A python package for automated ranking of classifiers. *J. Open Source Softw.* **2020**, *5*, 2173.
(48) Demšar, J. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **2006**, *7*, 1−30.

https://doi.org/10.1021/acsomega.4c02147
ACS Omega 2024, 9, 23940−23948