Artificial Intelligence in the Life Sciences 4 (2023) 100086

Contents lists available at ScienceDirect

Artificial Intelligence in the Life Sciences

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

Review

Exploring new horizons: Empowering computer-assisted drug design with few-shot learning

Sabrina Silva-Mendonça [a,b], Arthur Ricardo de Sousa Vitória [b,c], Telma Woerle de Lima [b,c], Arlindo Rodrigues Galvão-Filho [b,c], Carolina Horta Andrade [a,b,*]

[a] Laboratory for Molecular Modeling and Drug Design (LabMol), Faculty of Pharmacy, Universidade Federal de Goiás, Goiânia, GO 74605-170, Brazil
[b] Center for Excellence in Artificial Intelligence (CEIA), Institute of Informatics, Universidade Federal de Goiás, Goiânia, GO 74605-170, Brazil
[c] Institute of Informatics, Universidade Federal de Goiás, Goiânia, Brazil

ARTICLE INFO

Keywords:
Drug discovery
Low data
Few-shot learning
Meta-learning
Neglected diseases
Rare diseases

ABSTRACT

Computational approaches have revolutionized the field of drug discovery, collectively known as Computer-Assisted Drug Design (CADD). Advancements in computing power, data generation, digitalization, and artificial intelligence (AI) techniques have played a crucial role in the rise of CADD. These approaches offer numerous benefits, enabling the analysis and interpretation of vast amounts of data from diverse sources, such as genomics, structural information, and clinical trials data. By integrating and analyzing these multiple data sources, researchers can efficiently identify potential drug targets and develop new drug candidates. Among the AI techniques, machine learning (ML) and deep learning (DL) have shown tremendous promise in drug discovery. ML and DL models can effectively utilize experimental data to accurately predict the efficacy and safety of drug candidates. However, despite these advancements, certain areas in drug discovery face data scarcity, particularly in neglected, rare, and emerging viral diseases. Few-shot learning (FSL) is an emerging approach that addresses the challenge of limited data in drug discovery. FSL enables ML models to learn from a small number of examples of a new task, achieving commendable performance by leveraging knowledge learned from related datasets or prior information. It often involves meta-learning, which trains a model to learn how to learn from few data. This ability to quickly adapt to new tasks with low data circumvents the need for extensive training on large datasets. By enabling efficient learning from a small amount of data, few-shot learning has the potential to accelerate the drug discovery process and enhance the success rate of drug development. In this review, we introduce the concept of few-shot learning and its application in drug discovery. Furthermore, we demonstrate the valuable application of few-shot learning in the identification of new drug targets, accurate prediction of drug efficacy, and the design of novel compounds possessing desired biological properties. This comprehensive review draws upon numerous papers from the literature to provide extensive insights into the effectiveness and potential of few-shot learning in these critical areas of drug discovery and development.

Introduction

The field of Computer-Assisted Drug Design (CADD) has seen remarkable progress and is constantly evolving [1]. CADD techniques can be broadly classified as Structure-Based Drug Design (SBDD) [2] and Ligand-Based Drug Design (LBDD) [3], based on the information and availability of protein/target and ligand structures, respectively. The term artificial intelligence was introduced by John McCarthy in 1956, who defined it as the field dedicated to the creation of intelligent machines [4]. Essentially, AI encompasses the capacity of machines to carry out tasks in response to different conditions. In recent years, AI has garnered considerable attention for its aptitude to learn from data and accomplish specific tasks, and the pharmaceutical industry has equally recognized and embraced its immense potential [5–8]. AI algorithms are now widely applied in various computational approaches for CADD, including the prediction of 3D protein structures [9,10], the development of docking scoring functions [11], the execution of molecular docking [12,13], *de novo* design [14], the establishment of QSAR (Quantitative Structure-Activity Relationship) models [15], and the prediction of synthetic routes and synthetic accessibility [16], among

* Corresponding author.
E-mail address: carolina@ufg.br (C.H. Andrade).

https://doi.org/10.1016/j.ailsci.2023.100086
Received 30 June 2023; Received in revised form 29 August 2023; Accepted 4 September 2023
Available online 9 September 2023

others.

However, within the field of drug discovery, certain areas like neglected, tropical [17–20] and rare diseases [21–23] frequently face a scarcity of comprehensive datasets. This constraint gives rise to concern since models encounter difficulties in achieving high performance and effectively generalizing to compounds beyond the known chemical space [24]. This difficulty poses a significant obstacle for drug discovery, particularly when endeavoring to identify novel scaffolds from predicted compounds within academic and open science research.

Few-shot learning (FSL) is a specific subfield within ML that is utilized in scenarios where data resources are limited. It aims to achieve satisfactory learning performance despite the constraints of limited supervised information available in the training set. This training set consists of input examples along with their corresponding output labels [25]. Consequently, various techniques have been developed in order to enhance FSL, including transfer learning [26,27], neural network, and meta-learning [28,29]. These methods aim to address the limitations associated with learning from limited examples and improve the performance of models in FSL scenarios.

The Transfer Learning (TL) involves leveraging pre-existing knowledge from "upstream" tasks and applying it to a smaller, distinct but correlated, task, thereby enhancing the modeling performance even with limited samples [30]. In contrast, meta-learning in neural networks, often referred to as “learning to learn”, refers to the ability of an algorithm to improve its performance for each task through experience, considering the number of tasks as well [31]. This approach is applied to learning episodes obtained from a specific task family, resulting in a base learning algorithm that exhibits strong performance when faced with new tasks drawn from the same family [32].

These algorithms can be employed using various approaches to address the challenges posed by limited data. In this review, we show the application of FSL in the field of drug discovery, providing examples using QSAR and other methodologies. By highlighting these examples, we hope that researchers will be encouraged to explore and utilize FSL for modeling datasets with scarce resources.

Few-Shot learning

Few-Shot Learning (FSL) can be employed in various types of data for supervised learning problems. These include: (*i*) classification: this type of FSL approach, known as *N*-way-*K*-shot, involves considering *N* from the number of classes, and *K* is the number of examples. Alternatively, one-shot learning (OSL) is employed when only one example is available, while zero-shot learning (ZSL) is applied when zero examples are provided [33,34]; (*ii*) regression: FSL can also be applied to regression tasks, where the goal is to estimate a function based on only a few input-output example pairs sampled from that function. In this scenario, the output represents the observed value of the dependent variable, while the input records the observed value of the independent variable [25,35]. Furthermore, FSL can be used in semi- or unsupervised learning, using the reinforcement learning to weight orient the prior knowledge and prediction between the tasks [36,37].

The taxonomy of FSL revolves around the concept of an unreliable empirical risk minimizer [25]. In this framework, the hypothesis h is minimized based on the empirical risk, which is derived from the error decomposition of the model’s poor decision-making. When a sufficient number of supervised samples are available, the empirical risk can approximate to the expected risk, leading to favorable performance and accurate predictions [38,39].

Various methodologies have been developed to address this problem, and they can be categorized into different approaches to use prior knowledge. According with Wang and co-workers [25], based on which aspect is enhanced using prior knowledge, existing FSL works can be categorized into the following perspectives: (i) *Data*, where prior knowledge is used to augment and increase the number of samples, leveraging related, labeled, and unlabeled data can be employed to augment the low data available; (ii) *Model*, where prior knowledge is used to constrain the model and result in a smaller hypothesis space between empirical and expected risk, training the model on tasks similar to the target problem, it can acquire transferable knowledge and improve its generalization capabilities; and (iii) *Algorithm*, where prior knowledge alters the search strategy to better approximate empirical and expected risk, employing algorithms specifically designed for FSL can further refine predictions or develop new algorithms capable of generalizing predictions with minimal data (Fig. 1).

Apart from these, other authors presented alternative taxonomies for FSL, including metric-based, data-based, and optimizer-based approaches [40–42]. While some authors limit the scope of FSL to transfer learning and meta-learning, it is important to note that FSL is a broader problem and can be addressed using various approaches. It is worth mentioning that even automated machine learning (AutoML) [29] techniques can be adapted to tackle FSL problems.

In addition to the mentioned architectures, there are various other approaches that can be utilized for FSL, including Siamese networks [43], prototypical networks [44], kernel networks [45], bi-directional LSTM [46], convolutional neural networks [24], generative neural networks [47], matching networks [41], and others. The method of choice depends on the specific requirements and objectives of the problem, aiming to achieve accurate predictions even when only a few or zero examples are available.

We performed a search on Google Scholar and PubMed, using the keywords "Few-shot learning for CADD", "Few-shot learning for Drug Discovery", "Few-shot learning for Drug Design," "Meta-learning for CADD," "Meta-learning for Drug Discovery," "Low-resource for Drug Discovery," and "Meta-learning for Drug Design". These papers were classified based on their architectures and objectives, following the taxonomy proposed by Wang et al. [25], which includes data, model, and algorithm taxonomies. Some selected papers will be thoroughly discussed within its respective taxonomy category. Table 1 includes details of each paper found, including the dataset, features, architectures, type of prediction, taxonomy, metrics, and year of publication.

Recent advances in FSL for drug discovery

In the following sections, we will present and discuss the recent advances in FSF for drug discovery, exploring its three major branches: data, model, and algorithms.

Data

Within each category, there are various methodologies employed to tackle the FSL problem. Data augmentation can be achieved by utilizing the training set samples, weakly or unlabeled data sets, or similar data sets.

The three methodologies differ in terms of the type of data utilized in the training set. In the case of augmenting data using the sample of the training set, additional samples are generated by applying variations to the existing data. When augmenting data using weakly or unlabeled data sets, the augmentation process involves incorporating weakly or unlabeled samples from large data sets. Lastly, for augmenting samples using similar data sets, aggregation techniques are employed to combine similar samples from large data sets, which are commonly utilized in generative neural networks, in this approach, the large data should be labeled, different from weakly or unlabeled data [25].

The work described by Meng, Zhao, and King (2023) adapt to the taxonomy of transforming data from the training set, when the model performs the augmentation during modeling [81]. The authors used a motif-based task augmentation, a meta-learning framework that learns to adapt to new tasks with few examples by augmenting the training set with molecular motifs. This approach begins with recursive decomposition of molecules in the dataset into substructures until each substructure appears more frequently than a specified threshold. An

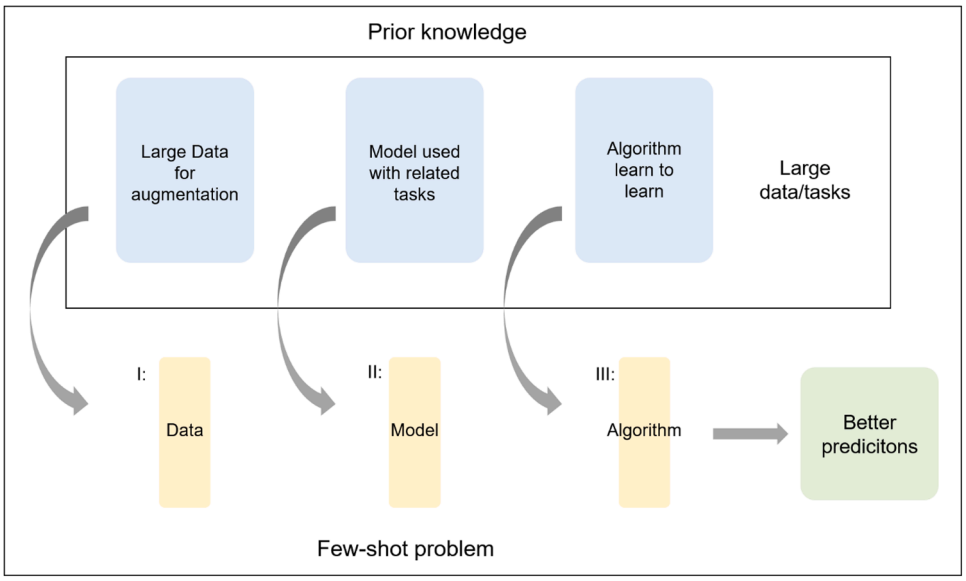

Fig. 1. Three key approaches to improve predictions for few-shot problems. Firstly, low data can be supplemented by incorporating related, labeled, and unlabeled data. Secondly, to enable generalization with only a few examples, the model can leverage insights from related tasks. Lastly, an algorithm can be refined or developed to enhance prediction generalization for scenarios with limited or no samples.

extracted motif vocabulary serves as external memory. In meta-training tasks, support/query samples retrieve relevant motifs from the vocabulary. Using convex combinations, the approach creates prototype clusters to represent each class. Then, the authors evaluated the effectiveness of their approach on several benchmark datasets for molecular property prediction. The results demonstrated that incorporating motif-based task augmentation significantly improves the performance of FSL models compared to conventional approaches [81]. Furthermore, the technique exhibits robustness across different molecular property prediction tasks, highlighting its potential for practical applications in drug discovery and other molecular-related domains.

The paper described by Gao and co-workers [63] can be classified as a framework to transform samples from similar datasets, when is aggregated or adapting data from similar and large data set. The authors compose a two-step approach where the low dimensional deep learning model captures the overall atomic properties, while the fragment-based GCN focuses on the local environment of atoms. In the first step, the low dimensional deep learning model is trained to predict atomic properties based on simplified atomic descriptors. This model provides an initial estimation of the atomic properties in a fast and efficient manner. In the second step, a fragment-based GCN is applied to refine the predictions by considering the interactions between atoms and their local environments [63]. Finally, the authors evaluated the performance of their method based on various datasets for atomic and inter-atomic property prediction. The results demonstrated that this combined approach achieved high accuracy while significantly reducing computational costs compared to traditional methods [63].

Model

The model concept is divided in four categories: multi-task learning, representation modeling, learning with external memory, and generative models. The initial classification is multi-task learning, which involves parameter sharing or use the same parameters the across tasks to facilitate learning. Although this taxonomy is widely applied and well established in drug discovery [70,86–96], it has been already revised and discussed previously by other authors [97,98] and therefore will not be the focus of this review. In the following sections, we will discuss representation learning, learning with external memory, and generative modeling.

Representation learning

Representation learning in FSL refers to the process of learning a low-dimensional representation or embedding space for the input data, where similar samples are closer to each other and dissimilar samples are farther apart (Fig. 2). This embedding space enables effective generalization and inference even with limited labeled data [99].

One popular approach in representation learning is to use deep neural networks (DNNs), such as graph convolutional neural networks (GCNs), to learn the representations. These networks are trained on large-scale datasets to capture the inherent structure and patterns in the data. The learned representation can then be used for FSL tasks by leveraging the similarity or distance between samples in the embedding space [41,44].

Language Model and GCNs (LM-GVP) is a deep learning framework designed for protein property prediction created by Wang and collaborators [68], incorporating both sequence and structural information. The technique aims to overcome the limitations of existing methods by leveraging the power of language models and GCNs. The language model processes protein sequences to capture sequential patterns, while the GCNs utilizes structural information encoded as graphs to capture spatial relationships among each amino acid, specifying the task. The framework is trained on a large dataset and demonstrates impressive performance in various protein property prediction tasks. Its extensibility allows for the incorporation of additional features and information sources, making it versatile and adaptable to diverse prediction problems. LM-GVP represents a promising approach for accurate and comprehensive protein property prediction by leveraging both sequence and structural characteristics of proteins [68].

Zhang and colleagues [84] introduced a novel technique for few-shot drug synergy prediction using a prior-guided hypernetwork architecture. This approach addresses the challenges in discovering effective drug combinations and advancing precision medicine. The hypernetwork learns prior knowledge about drug synergy from the training data and generates task-specific weights for the main network, allowing the model to adapt to different combinations. The FSL approach trains the model on limited drug combination examples to handle scenarios with scarce data. By incorporating additional prior information, such as drug-target interactions and molecular properties, the model gains valuable guidance and improves prediction accuracy. The model's performance is evaluated on a validation set to optimize hyperparameters,

Table 1
Results from the search on the engines Google Scholar and PubMed, using the keywords "Few-shot learning for CADD," "Few-shot learning for Drug Discovery," "Few-shot learning for Drug Design," "Meta-learning for CADD," "Meta-learning for Drug Discovery," "Low-resource for Drug Discovery," and "Meta-learning for Drug Design," performed on 05/2023, encompassing papers that apply or develop frameworks for the integration of few-shot learning in the field of drug discovery.

Authors	Dataset	Features	Architecture	Prediction	Taxonomy	Metrics	Year
[48]	Tox21, MUV and SIDER	Fingerprint	Attention LSTM	Molecular property	Model	AUC-ROC Tox21 (0.823 ± 0.002) MUV (0.499 ± 0.053) SIDER (0.669 ± 0.007)	2017
[49]	ENZYME database, Kinase.com, KEGG database	ProtVec representation	LSTM and Recurrent Neural Network	Kinase-phosphosite associations	Model	Hit top 1% (21.52)	2020
[50]	DrugBank, GNBR, Hetionet, STRING, IntAct, and DGIdb databases	Text embeddings	Relacional GCN	Biological activity	Model	Hit top 1% (25.32)	2020
[51]	ChEMBL20 database	ECFP4 and molecular graphs	Message passing neural networks and Model Agnostic Meta-Learning	Biological activity	Algorithm	AUC-PR CHEMBL ID 1,738,019 (0.903 ± 0.127) CHEMBL ID 2,095,143 (0.539 ± 0.282)	2020
[52]	ChEMBL20 database	Molecular graphs	Gated GCN and Model Agnostic Meta-Learning	Biological activity	Algorithm	AUC-PR CHEMBL ID 918,058 (0.747 ± 0.076) CHEMBL ID 1,804,798 (0.369 ± 0.021)	2020
[53]	NELL-One, COVID19-One	KGC embeddings	Meta pattern learning framework	Built KGC dataset COVID19-One	Data	Hit top 1% (0.188)	2021
[54]	Tox21 and SIDER	Molecular Graphs and SMILES	Model agnostic machine learning and GCN	Molecular property	Algorithm	AUC-ROC Tox21 1-shot (0.77) SIDER 1-shot (0.74) Tox21 5-shot (0.78) SIDER 5-shot (0.75)	2021
[28]	ChEMBL27 database	Molecular Graphs	Model Agnostic Meta-learning	Biological activity	Model	AUC-PR All enzymes (0.206 ± 0.008)	2021
[55]	Antimicrobial Peptide Database (APD), PROSO II database, TULA-2 and SHP-2 protein library.	Peptide sequence	GCN	Peptide design	Algorithm	ACC The best: anti-MRSA (0.93 ±0.02) The worst: soluble (0.50 ±0.02)	2021
[56]	ChEMBL20 database	Molecular Graphs	Model Agnostic Meta-learning	Biological activity	Algorithm	Averaged success rate (49.39%)	2021
[57]	ChEMBL database	RDKit descriptors	Model Agnostic Meta-learning and recurrent neural network	Molecular property	Algorithm	Averaged accuracy SIDER (70.01 ± 0.86%) Tox21 (71.07 ± 0.91%) MUV (60.66 ± 1.09%) ToxCast (74.02 ± 1.57%)	2021
[58]	ZINC15 database, ChEMBL database, FDA-approved drugs, and Ellinger dataset	Molecular Graphs, and PubChem fingerprints	GCN and Multi-layer Perceptron	Biological activity	Model	AUC-ROC Jak1 8:1:1 ratio (0.99 ± 0.11) F1 Jak1 8:1:1 ratio (0.92 ± 0.80)	2021
[59]	Cancer Cell Line Encyclopedia (CCLE) project from the DepMap website	Molecular Graphs	Multi-layer perceptron and Model Agnostic Meta-learning	Biological activity	Algorithm	KU-55,933 resistance (third from top; $R = 0.54$)	2021
[60]	United States Patent and Trademark Office (USPTO) dataset	Weighted ECFP	Modern Holpfield Network	Reaction prediction	Algorithm	Top-k Accuracy (%) USPTO-lg TOP 1 (16.9) TOP 10 (42.2) TOP 100 (72.4)	2021
[61]	MoleculeNet	Molecular Graphs	Model Agnostic Meta-Learning and GCN	Molecular property	Model	AUC-ROC (%) 1-shot Tox21 (83.01 ± 0.09) SIDER (74.46 ± 0.29) MUV (66.94 ± 1.12) ToxCast (73.63 ± 1.00)	2021
[62]	miniImagenet, Tox21, SIDER, MUV, ToxCast	Embeddings	Bi-directional LSTM, Multi-layer perceptron, and GCN	Biological activity	Model	Accuracy ± 95% confidence interval (0.252)	2021
[63]	Alchemy library	Multiple-level molecular fragments, RDKit descriptors, and QM calculated descriptors	Fragment-based GCN	NMR 13C chemical shifts	Data	MAE (C—C): 1.0 kcal/Mol MAE (C—H): 1.6 kcal/mol MAE (O—H): 1.8 kcal/mol	2022
[64]	Database of Antimicrobial Activity and Structure of Peptides and NCBI	Feature vector	Multi-Layer Perceptron	Biological activity	Algorithm	AUC-ROC Between 0.6 to 0.96	2022
[65]	Tox21 and SIDER	Molecular graph	Model agnostic machine learning and GCN	Molecular property	Algorithm	AUC-ROC (%) 1-shot Tox21 (78.27) 1-shot SIDER (76.83)	2022

(*continued on next page*)

Table 1 (*continued*)

Authors	Dataset	Features	Architecture	Prediction	Taxonomy	Metrics	Year
						5-shot Tox21 (78.81) 5-shot SIDER (77.60)	
[66]	Drug Bank	SMILES	GCN	Drug-drug interaction	Model	Accuracy average 5-way 1-shot common events (0.8379 ± 0.0172) fewer events (0.8263 ± 0.0170) rare events (0.7123 ± 0.0172)	2022
[67]	StarPepDB, BIOPEP-UWM and others	Text Convolution Neural Network (TextCNN) backbone as feature vectors	GCN and Prototypical Nertwork	Biological activity	Model	ACC, AUC, and MCC The best: ACP (0.9381, 0.9754, and 0.8775) The worst: PSBP (0.7292, 0.8368, and 0.4620)	2022
[68]	Gene Ontology and TAPE Database	Text embeddings	Language models and GCN	Protein property	Model	Wilcoxon signed-rank test p-values for CC, BP, MF tasks: 5.31e−5; 3.15e−27; 2.01e−34.	2022
[69]	GPCRdb and Kinase inhibitors	Embedding	Siamese network, AttLSTM, IterRefLSTM	Drug excipients	Model	AUC Kinases +1/−1 IterRefLSTM (0.980 ± 0.004) GPCR +1/−1 AttLSTM (0.695 ± 0.199)	2022
[70]	Large Scale Comparison (LSC) dataset with ChEMBL20 database	Molecular graphs	GCN and Multi-layer perceptron	Molecular property	Model	AUC-ROC (%) ChEMBL 10 (78.35 ± 1.07) ChEMBL 50 (80.54 ± 1.02) ChEMBL 100 (81.15 ± 0.59)	2022
[71]	Cell Painting Dataset,	ECFP	Hopfield Neural Network and Multi-layer perceptron	Biological activity	Model	Top-k accuracy (%) TOP 1 (10.4) TOP 5 (21.3) TOP 10 (30.6)	2022
[72]	FS-Mol benchmark dataset	RDKit descriptors	Prototypical network and Self Normalizing Neural Network	Biological activity	Algorithm	AUC-PR All data (0.223 ± 0.011)	2023
[73]	DrugBank and TWOSIDES dataset	Feature vectors and Knowledge biomedical graphs	GCN	Drug-drug interaction	Algorithm	ACC and AUC DrugBank (0.7167 and 0.8240) Twosides (0.6285 and 0.6865)	2023
[74]	DrugComb Portal Data set	Weighted matrix	Large pre-trained language model	Drug-drug interaction	Model	ΔAUC-PR 0-shot The worst: Pancreas (0.033) The best: Endometrium (0.564)	2023
[75]	MoleculeNet and FS-Mol benchmark data set, and PubChem Bioassay	RDKit descriptors, molecular graph, and SMILES	Contrastive Language-Image Pre-training	Biological activity	Algorithm	AUC-ROC (%) FS-Mol dataset (69.26 ± 0.2)	2023
[76]	Tox21 and SIDER	RDKit descriptors and Molecular Graph	GCN and model-agnostic meta-learning	Molecular property	Algorithm	AUC-ROC 5-shot (%) Tox21 (75.55) SIDER (70.32)	2023
[77]	Tox21 and SIDER	Molecular Graph	GCN and Transformer	Molecular property	Algorithm	AUC-ROC 5-shot Tox21 (0.7628) SIDER (0.7195)	2023
[78]	MoleculeNet benchmark	Molecular Graph	Graph kernel, GCN, Multi-layer perceptron and Model Agnostic Meta-learning	Molecular property	Algorithm	AUC 1-shot (%) Tox21 (84.09 ± 0.20) SIDER (77.53 ± 0.41) MUV (68.76 ± 1.05) ToxCast (74.40 ± 0.82)	2023
[79]	Tox21, MUV, QM9, and SIDER	Molecular Graphs	Model Agnostic Meta-learning and GCN	Molecular activity	Algorithm	AUC-ROC 12 training tasks from Tox21 (0.8962) 1 test set SIDER (0.5264)	2023
[80]	ChEMBL database	SMILES, ECFP	Transformers	Biological activity	Data	AUC 1-shot HDM (75.89(+1.82)) HDAC (79.46(+5.75)) HMT (78.12(+4.89))	2023
[81]	MoleculeNet bechmark	Molecular Graphs	Model Agnostic Meta-learning	Molecular property	Data	AUC-ROC (%) 1-shot Tox21 (84.15 ± 0.60) SIDER (76.53 ± 0.94) MUV (70.75 ± 1.15) ToxCast (75.29 ± 0.92)	2023
[82]	ChEMBL database	SMILES and Molecular graphs	GCN and Reinforcement Learning	Drug Design	Model	AUC PpIC50 (0.73) PlogP (0.78)	2023
[83]	Tox21, MUV, DUD-E	Molecular graph	Random Forest and GCN	Biological activity	Model	AUC-PR +1/−1 siameseNet (0.198 ± 0.102)	2023

(*continued on next page*)

Table 1 (*continued*)

Authors	Dataset	Features	Architecture	Prediction	Taxonomy	Metrics	Year
						matchingNet (0.352 ± 0.121) prototypicalNet (0.373 ± 0.102) relationNet (0.342 ± 0.093)	
[84]	SYNERGxDB database	Molecular graphs	Hypernetwork, Model Agnostic Meta-learning, and GCN	Drug-drug interaction	Model	5-shot MSE (0.115 ± 0.002) SCC (0.508 ± 0.004) R2 (0.180 ± 0.008)	2023
[85]	Buchwald-Hartwig HTE dataset, Suzuki-Miyaura HTE dataset and Buchwald-Hartwig ELN dataset	Density Functional Theory (DFT) descriptor	Model Agnostic Meta-Learning	Reaction prediction	Algorithm	Suzuki-Miyaura HTE Dataset RSME (18.3102) MAE (0.6350) R2 (13.8344)	2023

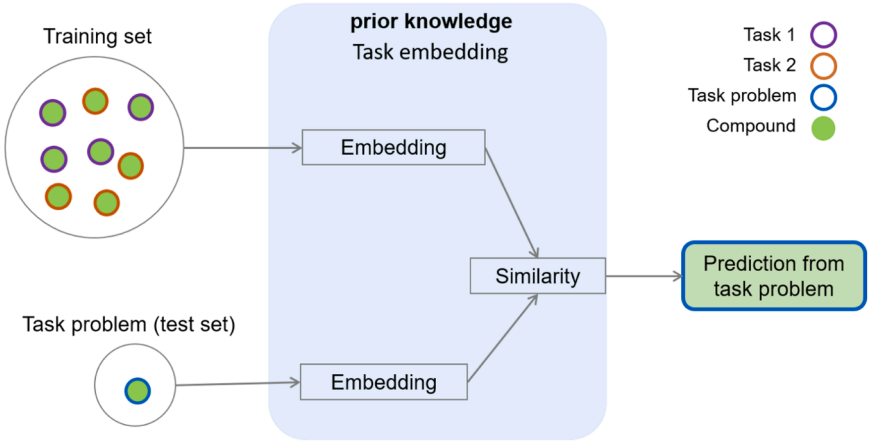

Fig. 2. The representation modeling concept. The tasks from prior knowledge and the task problem (test set) are embedded, and the similarity between the embeddings is calculated according to the features shared between them in order to accomplish the desired prediction.

and its effectiveness is assessed on a test set for drug synergy prediction. Overall, this approach shows promise in efficiently predicting drug synergies and advancing precision medicine.

Liu and co-workers [58] proposed a new methodology to identify potential drug candidates that can target multiple proteins associated with the SARS-CoV-2 virus. The researchers proposed a novel GCN architecture that effectively captures the structural and functional information of both drugs and proteins. The GCN model learns to encode the molecular structures and interactions into a low-dimensional embedding space, enabling efficient similarity calculations and predictions. This approach provides a valuable tool for identify multi-targeted drug candidates [58].

The paper described by Sanchez-Fernandez and co-workers [71] is another task-invariant example, which introduces CLOOME, a novel technique that utilizes contrastive learning to facilitate the querying of bioimaging databases using chemical structures. While bioimaging data provides valuable insights for drug discovery and biological processes, the lack of direct searchability based on chemical structure information poses a challenge. CLOOME addresses this issue by employing contrastive learning, a self-supervised learning approach, to establish associations between chemical structures and corresponding bioimages. The neural network is trained to maximize similarity between paired instances of the same chemical structure and bioimage while minimizing similarity between different pairs. This enables CLOOME to learn a joint embedding space where chemical structures and bioimages are semantically linked. By integrating chemical structure information with bioimaging data, CLOOME offers new opportunities for efficient and targeted analysis, allowing researchers to leverage large-scale bioimaging databases for drug discovery and biological research [71].

A closely related approach to QSAR is MetaHMEI, described by Lu and collaborators (2023) [80], which utilizes a meta-learning framework to predict few-shot modifications of histone modifying enzyme inhibitors (HMEIs). HMEIs play a crucial role in epigenetic regulation and have potential therapeutic applications. However, predicting effective modifications of HMEIs with few data is challenging. MetaHMEI addresses this challenge by leveraging meta-learning, which enables the model to learn from a diverse set of related tasks and generalize to new tasks with few labeled examples. The proposed model combines a Siamese network with a memory-augmented neural network, allowing it to capture similarities and differences between HMEIs and effectively predict modifications. The results on benchmark datasets demonstrate that MetaHMEI outperforms baseline methods in predicting few-shot modifications of HMEIs. The findings suggested that meta-learning can enhance the prediction accuracy and generalization capability in the context of modifying enzyme inhibitors, facilitating the development of novel therapeutic strategies targeting epigenetic regulation [80].

Wang, Yao and Dou [61] presented an innovative approach to enhance FSL for molecular property prediction. The technique employs property-aware relation networks to effectively capture and model the intricate relationships between different molecular properties. By focusing on specific properties of interest, the model demonstrates the ability to generalize and make accurate predictions even when confronted with restricted labeled data. The approach combines a GCN architecture with a property-aware relation module, enabling the model to capture and exploit property-specific interactions among molecules. Experimental evaluations validate the superior performance of this method compared to existing approaches in various few-shot molecular property prediction tasks. This advancement has significant implications for drug discovery, as it facilitates accurate predictions of molecular properties using limited data, ultimately reducing the time and

resources required for validation measurements.

Learning with external memory

Learning with external memory is a subcategory within the FSL taxonomy, wherein an external memory component is incorporated as an additional source of supplementary information to enrich the learning and inference process of the model. In this approach, the external memory serves as an auxiliary storage module that can be accessed for reading and writing during both training and inference stages (Fig. 3).

In the domain of drug discovery, although there are already published papers exploring the concept of fine-tuning parameters within the learning with external memory approach, the exploration of refining representations remains relatively unexplored in the existing literature. Seidl and co-workers [60] introduced a technique called Modern Hopfield Networks (MHopNets) to improve the prediction of reaction templates in FSL and ZSL scenarios. Reaction templates are essential for synthesizing new chemical compounds, but obtaining labeled data for all possible reactions is impractical. MHopNets leverage the power of neural networks and memory mechanisms to address this challenge. The approach involves training a MHopNets to store and retrieve reaction templates. By using an attention mechanism, the network learns to attend to relevant parts of the input and produce accurate predictions. The results demonstrated that MHopNets outperform existing methods in both FSL and ZSL settings, effectively predicting reaction templates even with limited labeled data. This technique has the potential to advance the field of chemical synthesis by enabling the discovery of novel reactions and facilitating the design of new compounds [60].

Chen and co-workers [85] presented a novel technique called MetaRF, which combines attention mechanisms with random forest models for predicting reaction yields with limited data. Accurate prediction of reaction yields is crucial for guiding chemical synthesis processes, but obtaining sufficient labeled data for training can be challenging. MetaRF addresses this problem by leveraging the power of attention mechanisms to selectively focus on relevant features in the input data. The approach involves training an attention-based random forest model that learns to assign importance weights to different features based on their relevance to the prediction task. The results demonstrated that MetaRF outperforms traditional random forest models and other state-of-the-art methods in predicting reaction yields with only a few trial data points. The technique shows promise in enabling more accurate and efficient reaction yield prediction, leading to improved chemical synthesis processes and accelerated drug discovery efforts [85].

Jiang and Gao [53] presented the MetaP framework, which consists of two main components: pattern mining and pattern matching. In the pattern mining phase, the meta pattern representation is used to capture the underlying structure of knowledge graphs. These meta patterns are learned from the existing graph data and aim to provide a compact yet informative way to represent graph patterns. The pattern mining employs meta-learning techniques, where the model is trained on multiple "meta-tasks." Each meta-task involves a small support set of known facts and a query set with missing facts. During the testing phase (pattern matching), MetaP is given a new, unseen knowledge graph completion task with only a few observed facts. Using the learned meta patterns, the model generalizes and predicts the missing facts for the given task. MetaP can utilize external information, such as embeddings or linguistic features, to enhance its knowledge graph completion performance. To evaluate the effectiveness of MetaP, the authors conducted experiments on benchmark datasets. The results demonstrated that MetaP outperforms existing state-of-the-art methods in terms of accuracy and efficiency, even with very limited labeled data [53].

Generative modeling

Incorporating generative models into FSL involves the generation of new data points or samples for a specific task using a limited number of labeled examples. These generative models are designed to capture the underlying distribution of the training data and generate novel samples that exhibit similarities to the existing data. By training these models on a small set of labeled examples, they can subsequently generate additional samples that serve to augment the available data for the target task.

An example of generative modeling is shown by Dong and coworkers [82], who introduced a novel approach to address the challenge of limited data in drug discovery using reinforcement learning and one-shot learning techniques to enable effective molecular design in low-data situations. The reinforcement learning agent explores a chemical space and iteratively generates molecules, maximizing a reward signal associated with the desired properties (specific tasks). Additionally, one-shot learning allows the model to make accurate predictions using minimal training examples, leveraging knowledge from labeled molecules to optimize properties of unseen molecules. The technique is evaluated in drug design tasks, demonstrating its effectiveness in generating high-quality molecules even with low data [82]. This approach represents a promising advancement in drug discovery, addressing the challenges of low-data situations and enabling efficient molecular design for improved drug candidates.

Other example is CancerGPT, from Li and collaborators [74], which utilizes the capabilities of pre-trained language models, specifically GPT

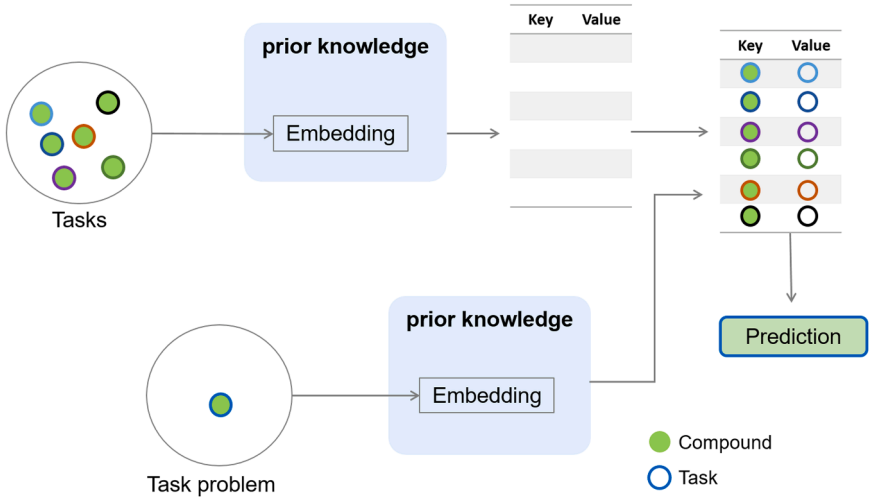

Fig. 3. The concept of learning with external memory. The prior knowledge of training set is stored in an external memory and weighting the test set based on this external memory. By combining the most similar embeddings, the model can make accurate predictions.

(Generative Pre-trained Transformer). These language models have undergone extensive training on large volumes of text data, enabling them to capture rich contextual information and semantic understanding. Through the process of fine-tuning the GPT model using a labeled dataset comprising drug pairs and their corresponding synergy scores, CancerGPT adeptly captures the complex interplay between drugs and their synergistic effects. Recognizing the limitations posed by scarce labeled data, the research paper introduces a FSL framework, empowering CancerGPT to generalize its understanding and make precise predictions for previously unseen drug combinations. Moreover, experimental evaluations on various cancer cell line datasets demonstrated that CancerGPT surpasses existing methods in predicting drug pair synergy with FSL [74]. The approach showcases remarkable accuracy and generalization capabilities, facilitating the identification of synergistic drug combinations even with limited labeled examples. These results underscore the potential of large pre-trained language models, such as GPT, in the domain of drug synergy prediction. By leveraging language models and FSL, CancerGPT offers a promising avenue for guiding cancer treatment decisions and expediting the discovery of effective personalized medicine combinations.

Algorithm

Refining existing parameters

This approach focuses on the fine-tuning of learned representations to better align with the specific characteristics of a new task. By iteratively refining the existing parameters, the model can enhance its performance and exhibit improved generalization when faced with new classes or tasks that have limited labeled examples (Fig. 4). The refinement process commonly employs gradient-based optimization techniques, such as stochastic gradient descent (SGD) or its variants, to iteratively update the model's parameters and optimize its performance for the target task [41,100]. Fine-tuning existing parameters by regularization in FSL refers to the process of refining the pre-trained parameters of a model using regularization techniques to prevent overfitting and improve generalization on new tasks with limited labeled data [101] (Fig. 4). The aggregate a set of parameters is the process of combining the parameters learned from multiple tasks or data instances to obtain a single set of parameters that can be used for generalization and inference [100].

Gull and Minhas [64] proposed a technique for the accurate prediction of antimicrobial peptides (AMPs) in different species using zero and FSL. AMPs are important components of the innate immune system and exhibit potential therapeutic properties. However, the diversity of AMPs across various species poses a challenge for accurate prediction. The technique, called AMP0, leverages a deep learning framework that combines ZSL and FSL approaches. In the ZSL phase, the model is trained on a large dataset containing labeled AMPs from one species and learns to generalize to unseen species using semantic embeddings. In the FSL phase, the model fine-tunes its parameters using a small set of labeled examples from the target species, enabling species-specific prediction. The model effectively captures the inter-species similarities and species-specific characteristics of AMPs [64]. This approach provides a valuable tool for species-specific antimicrobial peptides prediction, facilitating the discovery of novel antimicrobial agents and aiding in the development of new therapeutic strategies against infectious diseases.

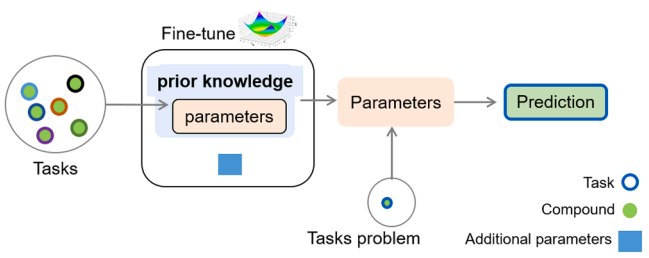

Fig. 4. The concept of refining existing parameters. Prior knowledge is used to learn how to refine and use additional parameters to few-shot, gradually, according to the necessity of additional parameters from the training set.

META-DDIE is an approach proposed by Den and co-workers [66] that addresses the challenging task of predicting drug-drug interaction events using FSL. Drug-drug interactions play a crucial role in drug safety and efficacy, and accurately predicting these interactions is essential for patient well-being. The proposed technique leverages a meta-learning framework combined with graph convolutional networks to enable effective learning from limited labeled examples. By training on a large dataset of drug-drug interaction networks, META-DDIE learns to generalize from a few labeled drug pairs and make accurate predictions for unseen drug interactions. This approach incorporates a graph-based representation of drugs, capturing their structural and functional properties, and employs attention mechanisms to focus on informative parts of the drug interaction networks. Experimental evaluation demonstrated that META-DDIE outperforms existing methods in predicting drug-drug interaction events, even with limited labeled data. The results highlight the potential of FSL and GCN in the domain of drug-drug interaction prediction, offering valuable insights for drug safety assessment and facilitating informed decision-making in healthcare.

Seidl and collaborators [75] integrated natural language processing (NLP) with activity prediction models to improve drug discovery process. Traditional activity prediction models rely solely on molecular features and structural information, often lacking the ability to interpret and utilize the wealth of information present in scientific literature. This paper introduced a novel framework that incorporates NLP techniques to extract relevant knowledge from textual sources, such as scientific papers and databases. By training the model to understand human language and extract key information, it can effectively augment the prediction models with additional context and insights. The proposed approach demonstrated significant improvements in activity prediction accuracy, outperforming conventional models that solely rely on molecular features [75]. The integration of NLP with activity prediction models opens new avenues for leveraging the vast amount of information available in scientific literature, enabling more informed decision-making in drug discovery and accelerating the identification of potential drug candidates.

Refining meta-learner parameters

Refining meta-learner parameters use the process of fine-tuning the parameters of the meta-learner model based on the performance and feedback received during the FSL tasks (Fig. 5). This step is crucial for improving the adaptation and generalization capabilities of the meta-learner model [100].

Wang and co-workers [56] presented a novel approach that tackles the challenge of optimizing molecular properties with limited data. Conventional optimization methods necessitate ample labeled data for training, which is typically lacking within molecular design. This study introduced a meta-learning framework that effectively utilizes prior knowledge from analogous molecules to facilitate the optimization process. By assimilating insights from a select group of high-quality molecules with known properties, the model exhibits the ability to generalize and provide accurate predictions for previously unseen molecules. To accomplish this, the approach combines a GCNs with a meta-learner, allowing the model to effectively capture and exploit the structural and property relationships among different molecules. Experimental evaluation demonstrated that this approach achieved superior performance compared to baseline methods, effectively optimizing molecular properties even with limited data [56]. This technique has significant implications for drug discovery and material design, as it enables efficient and effective optimization of molecular properties with minimal labeled examples, reducing the time and resources required for experimental measurements.

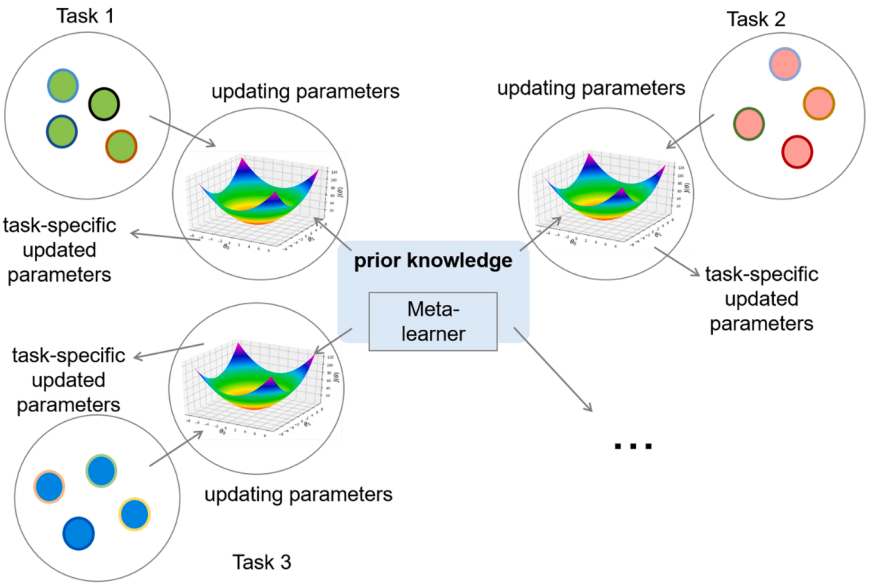

Fig. 5. The concept of refining meta-learner parameters. The gradient descent is used to refine the parameters of each task, then the meta-learner learns with the new parameters the refine again, using the gradient descent for the new tasks each appears, until find the best generalizing parameter refined using the tasks applied in the meta-learner.

Pappu and Paige [51] suggested an additional noteworthy example that harnesses the power of GNNs to effectively model the intricate relationships and structures of molecules. However, the efficacy of GNNs is often hindered by the scarcity of labeled data. To overcome this limitation, this study proposed an innovative framework that enhances the performance of GNNs in low-data scenarios. The technique integrates a blend of semi-supervised learning and self-supervised learning, utilizing both labeled and unlabeled data to enhance the model's ability to generalize. Moreover, the authors introduce a self-supervised pretext task that incentivizes the GNN to capture crucial features and patterns within the data [51]. The technique has the potential to significantly advance molecular machine learning by enabling accurate predictions and insights even with limited labeled data, paving the way for accelerated drug discovery and material design.

Lv and co-workers [73] employed drug-drug interaction prediction as a means to comprehend potential adverse effects and optimize drug combination therapies. Their approach capitalizes on a 3D GCNs, which effectively captures both the structural and spatial information of drugs. Additionally, the incorporation of FSL enables the model to generalize effectively from a limited number of labeled drug pairs, enabling accurate predictions for unseen pairs. Evaluation conducted on benchmark datasets demonstrate that the proposed model surpasses state-of-the-art methods when predicting drug-drug interactions in scaffold-based cold start scenarios [73]. These results underscore the potential of the 3D GCNs, combined with FSL, as a potent strategy for improving predictions in scenarios where data availability is limited. Consequently, this approach facilitates the identification of potential drug interactions and significantly contributes to drug discovery endeavors.

Ma and collaborators [59] developed predictive models for drug response that demonstrate a remarkable ability to translate from high-throughput screens to individual patients. This technique effectively tackles the challenge of limited labeled data by leveraging knowledge from a small set of examples and generalizing it to predict drug response in novel contexts. The researchers trained the FSL model on the combined dataset of high-throughput screens and individual patient profiles. The model learned to extract essential features from the data and to generalize drug response patterns across different biological contexts. A critical aspect of the study was to determine if the predictive models developed from high-throughput screens could be effectively transferred to predict drug responses for individual patients. The researchers tested the model's performance on patient-specific data, aiming to demonstrate its potential in enabling personalized medicine. The paper includes an in-depth analysis of the model's predictions and offers insights into the features and factors driving drug responses in both high-throughput screens and individual patients. Understanding these underlying mechanisms is essential for advancing precision medicine and drug development. These findings underscore the immense potential of FSL in the area of personalized medicine, facilitating the development of predictive models that can guide informed drug selection and significantly improve patient outcomes.

Learning optimizer

Learning the optimizer in FSL refers to the process of training a model to learn an adaptive optimization algorithm that can effectively adapt to different FSL tasks. This approach aims to improve the generalization and performance of FSL models by dynamically adjusting their optimization process based on the specific task at hand (Fig. 6) [102].

Stanley and co-workers [28] presented FS-Mol, a comprehensive dataset specifically designed for FSL within the domain of molecular compounds and properties. This pioneering work aims to overcome the limitations of existing molecular datasets by offering a diverse collection of molecular structures and properties, along with associated task definitions. FS-Mol encompasses a wide range of molecular properties, including aqueous solubility, bioactivity, and toxicity, providing a holistic evaluation framework for assessing FSL models. The dataset empowers researchers to evaluate the performance of different algorithms and models within a FSL context, thereby fostering the development of more robust and effective approaches amongst molecular property prediction. The statistical results and benchmarks achieved on FS-Mol clearly demonstrate its significance in advancing the understanding and refinement of FSL methods for molecular compounds. As a result, subsequent papers have begun utilizing the FS-Mol dataset [72,75,103], enhancing both the architecture and performance of predictions while furthering research in this area.

Yao and collaborators [65] proposed an approach using chemical property relations to guide the FSL process. The model learns to encode and exploit the relationships between chemical properties, enabling it to make accurate predictions even with limited labeled data. The technique combines a Siamese neural network architecture with a relation network module, which captures and utilizes the chemical property relations

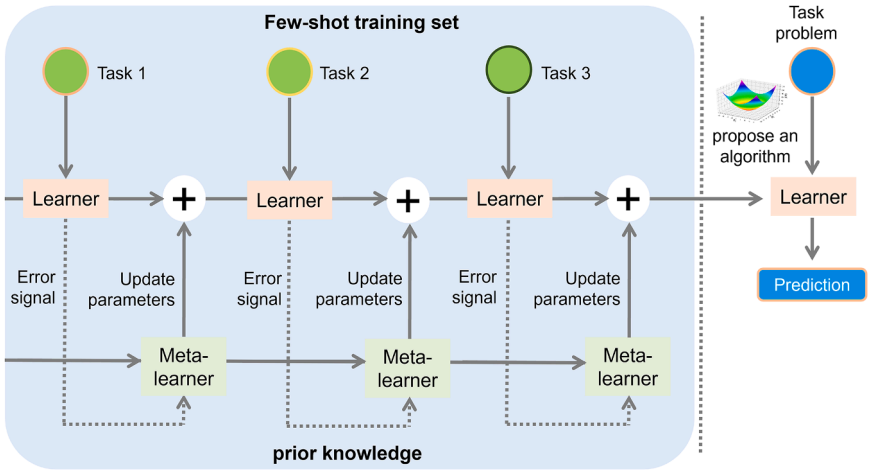

Fig. 6. The concept of learning optimizer. The prior knowledge is trained with related task, each task has an error signal and the meta-learner will update an optimizer with new parameters, according for each task. In the test set, the model will propose an algorithm with new generalizing parameters and optimizer, using the gradient descent to reduce the loss function and achieve the best algorithm to make good predictions.

effectively. Experimental evaluations on various molecular property prediction tasks demonstrate that the proposed method outperforms existing approaches, showcasing its effectiveness in handling few-shot scenarios. By leveraging the relationships between chemical properties, the technique enhances the model's ability to generalize and make reliable predictions with limited data.

In their research, He and co-workers [67] developed a novel technique that utilizes mutual information, a statistical dependence measure, to establish connections between peptide sequences and their bioactivity. This approach involves extracting informative features from the peptide sequences, including amino acid composition, physicochemical properties, and structural motifs, to identify potential bioactive peptides. They adopted a mutual information-based meta-learning framework to train a meta-model capable of adapting quickly to new peptide datasets. By learning from diverse datasets and corresponding predictive models, the meta-model demonstrated robust generalization and accurate predictions on previously unseen datasets. Data augmentation techniques were applied to generate synthetic peptide data, introducing variations to the existing sequences, thereby creating a more extensive and diverse dataset for training purposes. Upon training the meta-model, it exhibited the ability to predict peptide bioactivity from previously unseen datasets. The proposed methodology underwent thorough evaluation using benchmark datasets and comparison with state-of-the-art methods, showcasing its potential as a promising solution to accelerate bioactive peptide discovery.

By employing a meta-learning framework, the model learns to quickly adapt and make accurate predictions on new peptide sequences with limited labeled data. The approach was evaluated on multiple bioactive peptide datasets, demonstrating its effectiveness in accelerating the discovery of bioactive peptides compared to traditional methods. The results highlight the potential of mutual information-based meta-learning in enhancing the efficiency and effectiveness of bioactive peptide discovery, facilitating the development of novel therapeutics and biotechnological applications.

In another notable contribution, Lv and collaborators [79] introduced a technique for low-data drug discovery by employing meta learning in conjunction with graph attention networks. This approach addresses the challenge of limited labeled data by leveraging the power of meta learning, which enables the model to quickly adapt and learn from a few examples. Graph attention networks are utilized to effectively capture the structural information of molecules and learn their representations. The proposed method combines the meta learning framework with graph attention networks to train a model that can generalize well to unseen molecules and accurately predict their properties. Extensive experimental evaluations conducted on multiple drug discovery tasks convincingly demonstrate the effectiveness of this approach in low-data scenarios, surpassing the performance of traditional machine learning methods. These findings underscore the tremendous potential of meta learning with graph attention networks in accelerating the drug discovery process and facilitating the identification of potential candidates even with limited labeled data.

FSL models and their performance in property and biological activity prediction

Focusing on papers that applies FSL for molecular property prediction, we found around eight manuscripts which use toxicity data sets such as Tox21, MUV, SIDER, QM9 and ToxCast [48,54,61,65,76,78,79,81]. The papers listed above focus on few-shot molecular property prediction and adopt various techniques to address this challenging task. Notably, the best metrics were achieved by Meng and collaborators for Tox21, MUV, SIDER, and ToxCast [81], followed by Ju et al. [78] and Wang and colleagues [61]. Interestingly, Meng et al., adopted the architecture from Wang et al. [61], as a benchmark and applied augmented the data to them, resulting in noticeable improvements, especially for the challenging MUV dataset (AUC-ROC rose from 60.66% $\pm$ 1.09 to 70.75% $\pm$ 1.15). Despite the superior metrics achieved by Meng et al., the models' interpretability was not explicitly explained, unlike Wang et al., which demonstrated interpretability by capturing relationships between molecular properties and predictions.

In terms of architecture, Altae-Tran et al. [48] appears to employ a relatively simpler and more straightforward design compared to the other analyzed papers. Since the focus is on one-shot learning, their model is tailored to learn from a single instance and make predictions accordingly, potentially involving less complex computations and a reduced number of layers in comparison to other methods that incorporate advanced techniques such as graph attention networks or transformers. Conversely, papers by Meng et al. [81], Lv et al. [73], and Torres et al. [76] may feature more intricate architectures. These papers introduce meta-learning approaches or utilize advanced graph-based models and transformers, which typically require more sophisticated computations and a deeper network structure to effectively capture and process the complex relationships in molecular data. Despite that, Altae-Tran et al. [48] presented the simplest architecture, they don't achieve the same metrics for MUV and SIDER (see Table 1), and don't show the interpretability of their models.

On the other side, assessing the biological activity of molecules presents unique challenges, primarily due to the diverse targets utilized

to enhance their performance [28,50,72,75,80,83,51,52,56,58,59,62,64,67]. This diversity underscores the need for a standardized approach that can be employed in research endeavors seeking to leverage FSL methodologies. The work described by Stanley et al. [28] contributes significantly to this field by providing an accessible dataset, the FS-Mol, that facilitates the refinement and application of FSL techniques to biological activity prediction. However, only a limited number of studies, like Schimunek et al. [72], have utilized the FS-Mol dataset to enhance existing methodologies, demonstrating improved efficacy and accuracy in their predictions.

Challenges in FSL for CADD

Despite its promise, few-shot learning for CADD faces many challenges, such as the need for high-quality and diverse training data, model interpretability, and addressing issues of bias and fairness, especially when applying personalized medicine approaches.

High-quality and diverse training data

The quality of available data can vary widely, and bias can be introduced during data collection or curation. Biased or noisy data can lead to suboptimal model performance and potentially biased drug discovery outcomes. As well as data that span a wide range of molecular structures, targets, and properties. Ensuring that the -FSL model is exposed to diverse chemical and biological spaces is challenging. Limited diversity in the training data can lead to poor generalization. To exemplify this, Stanley and collaborators [28] used the ChEMBL data only with IC_{50} and/or EC_{50} and compounds with high molecular weight, is worth mention even with their encouraging results, ChEMBL data has different levels of confidence, which is crucial for any methodology of QSAR.

Interpretability

Understanding why a FSL model makes a particular prediction is crucial in drug design. Black-box models may not provide interpretable explanations, which can be problematic when trying to identify potential drug candidates or understand molecular interactions. Therefore, interpretability of FSL models is a significant challenge in the field. Among the FSL papers analyzed, the level of interpretability varies. Wang et al. [61] demonstrated the interpretability of their models by capturing the relationships between molecular properties and predictions. Lv et al. [73], Yao et al. [65], and Ju et al. [78] offered insights into crucial molecular interactions, chemical property relations, and improving interpretability. Lu et al. [80], Liu et al. [58], interpreted their generated FSL models with experimental results and experimental data already published in literature, and He et al. [67] presented the relationship between the peptides residues and the models learning, which are very important to understand the model and how they learn with the data. In contrast, Altae-Tran et al. [48], Guo et al. [54], and Torres et al. [76] lack interpretability of models due to the limitation or complexity of the FSL models.

Ethical issues and regulatory acceptance

As FSL become more influential in drug discovery, ethical concerns related to their use may arise. There could be questions about the fairness and transparency of these models, especially when they impact patient treatments and outcomes. Moreover, convincing regulatory agencies (such as FDA, EPA and others) to accept and validate the use of FSL models in drug discovery and toxicity research is a significant challenge. There may be a need to develop new regulatory frameworks and guidelines.

Model complexity

FSL models can use different levels of complexity such as augmenting data and learn an algorithm or augmenting data with representation learning. Due to the new subject of FSL for biological activity prediction, the use of different levels and types of data become difficult to compare the complexity between the models, with exception of manuscripts using the FS-Mol dataset [28,72]. On the other side, for molecular property prediction is possible to compare the approaches published by Meng et al. [81], Ju et al. [78], and Wang et al. [61] that used the same data (MUV, SIDER, Tox21, and ToxCast). This comparison revealed that the complexity of models can to be used to achieve a better performance for complex data.

Addressing these challenges will require interdisciplinary collaboration between machine learning experts, medicinal chemists, biologists, and regulatory agencies. Moreover, advances in data collection and curation, and model interpretability techniques specific to drug discovery will be crucial for the successful application of few-shot learning in this field.

Perspectives and guidelines for using FSL for CADD

As shown in this review, there are many successes cases emphasizing the significance of FSL in CADD, particularly when supported by compelling experimental results. Molecular property-related studies often rely on established datasets for model generation, facilitating effective comparisons with existing literature. In contrast, biological activity-focused manuscripts employ diverse datasets, such as ChEMBL, BindingDB, ZINC, FDA-approved drugs, Ellinger, and others, which hinders the establishment of a universal benchmark for model evaluation. One promising approach to kickstart this process involves utilizing the FS-Mol dataset, made available by Stanley et al. [28], enabling the exploration of novel models and advancing interpretability in AI-driven drug discovery—an area of ongoing importance.

The undeniable efficacy of FSL is evident through its impressive results and predictions. However, the prevalence of complex models raises the question of whether such intricacy is truly essential. Interestingly, Occam's Razor theory is often adhered to in model selection, favoring simplicity as the optimal approach [104,105]. Notably, certain targets exhibit strong predictive capabilities even without resorting to complex models [58]. These observations prompt the need for further investigation into the significance of meta-learning models and deep learning in FSL. While some molecular property datasets indicate improved metrics with more model complexity [61], the question remains less resolved for biological activity datasets, warranting more research to identify simple and explainable FSL models capable of surpassing the predictive performance of traditional machine and deep learning models.

As discussed, FSL can indeed be applied to various aspects of CADD. If one wants to start exploring FSL approaches, first need to select a specific CADD task you want to address with FSL, which can be predicting the activity of molecules against a specific target, identifying potential drug candidates, or optimizing molecular structures for specific properties. Then, you should gather relevant data, which is crucial for developing your models. Depending on the task, you may need molecular structure data, target protein data, bioactivity data, or other relevant information. It important to ensure that the data is clean, well-curated, and properly annotated. Then, you need to preprocess the data and extract relevant features. Next, you need to develop the FSL models, starting with simpler models and progressively exploring more complex ones. Evaluate the performance of your FSL models using appropriate metrics. Fine-tune hyperparameters and model architectures based on your evaluation results. As discussed, interpretability is crucial for CADD. Therefore, try to explore methods for interpreting FSL model decisions, and visualize results and model predictions to gain insights. Finally, experiment with different model architectures, loss functions,

and data representations to improve performance. Certainly, starting with a clear problem statement, relevant data, and a solid understanding of FSL principles will help make significant progress in applying FSL to CADD.

Conclusions

FSL has several potential applications in the field of drug design and discovery. As perspectives on how this approach could be used: 1. Identifying new drug targets: one potential application of FSL in drug discovery is to identify new drug targets based on few data. By training models to recognize patterns and features in small data sets related to different diseases, FSL can be used to identify promising new drug targets and accelerate drug discovery. 2. Predicting drug activity/efficacy: FSL could also be used to predict the activity of new drug candidates based on low data. To recognize common features and properties of effective drugs to predict which new compounds are likely to be effective based on their chemical properties and other characteristics. 3. Designing new drugs: Another potential application of FSL in drug discovery is to design new drugs based on limited data. By training models to recognize common features and properties of existing drugs to generate new drug candidates with similar properties, but optimized for specific targets or indications. 4. Personalized medicine: FSL could also be used to develop personalized medicine approaches based on restricted data from individual patients, to predict which treatments are likely to be most effective for individual patients based on their unique characteristics and medical histories. Therefore, we can conclude that FSL has the potential to accelerate drug discovery and improve personalized medicine by enabling models to learn and generalize to new tasks and domains. However, there are still several challenges that need to be addressed, such as data quality, model interpretability, and ethical considerations.

In this comprehensive review, we provide an in-depth analysis of the latest research papers in the field, focusing on the growing significance of representation learning approaches in accomplishing remarkable results for FSL tasks. The appeal of this approach lies in its straightforward calculation of similarities between embeddings, as opposed to the utilization of intricate meta-learning architectures, making it particularly intriguing. Nevertheless, delving deeper into the field of FSL with meta-learning models in the context of drug discovery, even with the demand for substantial computational resources, offers promising opportunities to tackle diverse challenging endpoints like ADME/Tox, as well as neglected, tropical, and rare diseases that may suffer from limited resources. While the introduction of new methodologies is undoubtedly valuable, there still exists untapped potential for exploring these methodologies in crucial areas such as scaffold generalization, explainable machine learning, and the chemical space encompassed by these models. Finally, bridging the gap between computational science and medicinal chemistry is crucial to develop models that not only perform well but also possess the capability to accurately and cost-effectively predict new drugs, solve protein structure-related problems, assess drug synergy probabilities, and evaluate drug toxicity and ADME properties. Such advancements may ultimately lead to a reduction in investments allocated to unreliable drugs.

Declaration of Competing Interest

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

Data availability

No data was used for the research described in the article.

Acknowledgments

Authors are grateful to the Brazilian funding agencies, CNPq (#440373/2022-0), CNPq BRICS STI COVID-19 (#441038/2020-4), FAPEG (#202010267000272) and CAPES (finance code 001). CHA is CNPq research fellow. TWL is CNPq technological development fellow.

References

[1] Brown FK, Sherer EC, Johnson SA, Holloway MK, Sherborne BS. The evolution of drug design at Merck Research Laboratories. J Comput Aided Mol Des 2017;31:255–66. https://doi.org/10.1007/s10822-016-9993-1.
[2] Sliwoski G, Kothiwale S, Meiler J, Lowe EW. Computational methods in drug discovery. Pharmacol Rev 2014;66:334–95. https://doi.org/10.1124/pr.112.007336.
[3] Prieto-Martínez FD, López-López E, Eurídice Juárez-Mercado K, Medina-Franco JL. Computational drug design methods—current and future perspectives. Silico Drug Des. 2019. https://doi.org/10.1016/b978-0-12-816125-8.00002-x.
[4] McCarthy J, Hayes PJ. Some philosophical problems from the standpoint of artificial intelligence. Readings artif. intell. Stanford: Elsevier; 1981. p. 431–50. https://doi.org/10.1016/B978-0-934613-03-3.50033-7.
[5] Mak KK, Pichika MR. Artificial intelligence in drug development: present status and future prospects. Drug Discov Today 2019;24:773–80. https://doi.org/10.1016/j.drudis.2018.11.014.
[6] Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. Drug Discov Today 2021;26:80–93. https://doi.org/10.1016/j.drudis.2020.10.010.
[7] Ramesh AN, Kambhampati C, Monson JRT, Drew PJ. Artificial intelligence in medicine. Ann R Coll Surg Engl 2004;86:334–8. https://doi.org/10.1308/147870804290.
[8] Bajorath J. Artificial intelligence in interdisciplinary life science and drug discovery research. Futur Sci OA 2022;8. https://doi.org/10.2144/fsoa-2022-0010.
[9] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596:583–9. https://doi.org/10.1038/s41586-021-03819-2.
[10] Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science (80-) 2021;373:871–6. https://doi.org/10.1126/science.abj8754.
[11] Wójcikowski M, Ballester PJ, Siedlecki P. Performance of machine-learning scoring functions in structure-based virtual screening. Sci Rep 2017;7:46710. https://doi.org/10.1038/srep46710.
[12] Ballester PJ, Mitchell JBO. A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. Bioinformatics 2010;26:1169–75. https://doi.org/10.1093/bioinformatics/btq112.
[13] Gentile F, Agrawal V, Hsing M, Ton AT, Ban F, Norinder U, et al. Deep docking: a deep learning platform for augmentation of structure based drug discovery. ACS Cent Sci 2020;6:939–49. https://doi.org/10.1021/acscentsci.0c00229.
[14] Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. Sci Adv 2018;4:eaap7885. https://doi.org/10.1126/sciadv.aap7885.
[15] Kumar V, Saha A, Roy K. Multi-target QSAR modeling for the identification of novel inhibitors against Alzheimer's disease. Chemom Intell Lab Syst 2023;233:104734. https://doi.org/10.1016/j.chemolab.2022.104734.
[16] Coley CW, Barzilay R, Jaakkola TS, Green WH, Jensen KF. Prediction of organic reaction outcomes using machine learning. ACS Cent Sci 2017;3:434–43. https://doi.org/10.1021/acscentsci.7b00064.
[17] Global report on neglected tropical diseases 2023. Geneva: world Health Organization: licence: CC BY-NC-SA 3.0 IGO; 2023.
[18] Toor J, Hamley JID, Fronterre C, Castaño MS, Chapman LAC, Coffeng LE, et al. Strengthening data collection for neglected tropical diseases: what data are needed for models to better inform tailored intervention programmes? PLoS Negl Trop Dis 2021;15:e0009351. https://doi.org/10.1371/journal.pntd.0009351.
[19] Alonso PL. Malaria: a problem to be solved and a time to be bold. Nat Med 2021;27:1506–9. https://doi.org/10.1038/s41591-021-01492-6.
[20] Phillips M.A., Burrows J.N., Manyando C., van Huijsduijnen R.H., Van Voorhis W.C., Wells TNC. Malaria. Nat Rev Dis Prim 2017;3:17050. https://doi.org/10.1038/nrdp.2017.50.
[21] Rare diseases, common challenges. Nat Genet 2022;54:215–. https://doi.org/10.1038/s41588-022-01037-8.
[22] Mitani AA, Haneuse S. Small data challenges of studying rare diseases. JAMA Netw Open 2020;3:e201965. https://doi.org/10.1001/jamanetworkopen.2020.1965.
[23] Angural A, Spolia A, Mahajan A, Verma V, Sharma A, Kumar P, et al. Review: understanding rare genetic diseases in low resource regions Like Jammu and Kashmir – India. Front Genet 2020;11. https://doi.org/10.3389/fgene.2020.00415.
[24] Garcia V, Bruna J. Few-shot learning with graph neural networks. Prepr ArXiv 2017. https://doi.org/10.48550/arXiv.1711.04043.
[25] Wang Y, Yao Q, Kwok JT, Ni LM. Generalizing from a Few Examples: a Survey on Few-shot Learning. ACM Comput Surv 2020;53:1–34. https://doi.org/10.1145/3386252.

[26] Han Y, Wang Z, Chen A, Ali I, Cai J, Ye S, et al. A deep transfer learning-based protocol accelerates full quantum mechanics calculation of protein. Brief Bioinform 2023;24. https://doi.org/10.1093/bib/bbac532.

[27] Wang X, Gao C, Han P, Li X, Chen W, Rodríguez Patón A, et al. PETrans: de novo drug design with protein-specific encoding based on transfer learning. Int J Mol Sci 2023;24:1146. https://doi.org/10.3390/ijms24021146.

[28] Stanley M., Bronskill J.F., Maziarz K., Misztela H., Lanini J., Segler M., et al. FS-Mol: a Few-Shot Learning Dataset of Molecules. 35th Conf Neural Inf Process Syst (NeurIPS 2021) Track Datasets Benchmarks 2021.

[29] Olier I, Sadawi N, Bickerton GR, Vanschoren J, Grosan C, Soldatova L, et al. Meta-QSAR: a large-scale application of meta-learning to drug design and discovery. Mach Learn 2018;107:285–311. https://doi.org/10.1007/s10994-017-5685-x.

[30] Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. J Big Data 2016;3:9. https://doi.org/10.1186/s40537-016-0043-6.

[31] Vilalta R, Drissi Y. A perspective view and survey of meta-learning. Artif Intell Rev 2002;18:77–95. https://doi.org/10.1023/A:1019956318069.

[32] Hospedales TM, Antoniou A, Micaelli P, Storkey AJ. Meta-learning in neural networks: a survey. IEEE Trans Pattern Anal Mach Intell 2021. https://doi.org/10.1109/TPAMI.2021.3079209. 1–1.

[33] Qiao S, Liu C, Shen W, Yuille A. Few-shot image recognition by predicting parameters from activations. Prepr ArXiv 2017. https://doi.org/10.48550/arXiv.1706.03466.

[34] Larochelle H, Erhan D, Bengio Y. Zero-data learning of new tasks. In: Proc AAAI 2008 Proc Twenty-Third AAAI Conf Artif Intell. 2; 2008. p. 646–51.

[35] Zell A., Sumbul G., Demir B. Deep Metric Learning-Based Semi-Supervised Regression With Alternate Learning 2022. https://doi.org/10.1109/ICIP46576.2022.9897939.

[36] Lake BM, Salakhutdinov R, Gross J, Tenenbaum JB. One shot learning of simple visual concepts. In: Proc 33rd Annu Meet Cogn Sci Soc CogSci 2011. 33; 2011. p. 2568–73.

[37] Wang Q, Chen K. Multi-label zero-shot human action recognition via joint latent ranking embedding. Neural Networks 2020;122:1–23. https://doi.org/10.1016/j.neunet.2019.09.029.

[38] Lake BM, Salakhutdinov R, Tnenbaum JB. Human-level concept learning through probabilistic program induction. Science (80-) 2015;350:1332–8. https://doi.org/10.1126/science.aab3050.

[39] Fei-Fei Li, Fergus R, Perona P. One-shot learning of object categories. IEEE Trans Pattern Anal Mach Intell 2006;28:594–611. https://doi.org/10.1109/TPAMI.2006.79.

[40] Yin W. Meta-learning for Few-shot Natural Language Processing: a Survey 2020. https://doi.org/2007.09604v1.

[41] Vinyals O., Blundell C., Lillicrap T., Kavukcuoglu K., Wierstra D. Matching Networks for One Shot Learning. Prepr ArXiv 2016. https://doi.org/10.48550/arXiv.1606.04080.

[42] Veríssimo GC, Serafim MSM, Kronenberger T, Ferreira RS, Honorio KM, Maltarollo VG. Designing drugs when there is low data availability: one-shot learning and other approaches to face the issues of a long-term concern. Expert Opin Drug Discov 2022;17:929–47. https://doi.org/10.1080/17460441.2022.2114451.

[43] Koch G, Zemel R, Salakhutdinov R. Siamese neural networks for one-shot image recognition. ICML Deep Learn. Work. 2015.

[44] Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning. Prepr ArXiv 2017:4078–88. https://doi.org/10.48550/arXiv.1703.05175. 2017-Decem.

[45] Przewięźlikowski M, Przybysz P, Tabor J, Zięba M, HyperMAML Spurek P. Few-shot adaptation of deep models with hypernetworks. Prepr ArXiv 2022. https://doi.org/10.48550/arXiv.2205.15745.

[46] Zhao J, Lin X, Zhou J, Yang J, He L, Yang Z. Knowledge-based fine-grained classification for few-shot learning. In: Proc. 2020 IEEE Int. Conf. Multimed. Expo. IEEE; 2020. p. 1–6. https://doi.org/10.1109/ICME46284.2020.9102809.

[47] Robb E, Chu WS, Kumar A, Huang JB. Few-shot adaptation of generative adversarial networks. Prepr Arxiv 2020. https://doi.org/10.48550/arXiv.2010.11943.

[48] Altae-Tran H, Ramsundar B, Pappu AS, Pande V. Low data drug discovery with one-shot learning. ACS Cent Sci 2017;3:283–93. https://doi.org/10.1021/acscentsci.6b00367.

[49] Deznabi I, Arabaci B, Koyuturk M, Tastan O. DeepKinZero: zero-shot learning for predicting kinase-phosphosite associations involving understudied kinases. Bioinformatics 2020;36:3652–61. https://doi.org/10.1093/bioinformatics/btaa013.

[50] Ioannidis VN, Zheng D, Karypis G. Few-shot link prediction via graph neural networks for Covid-19 drug-repurposing. Prepr ArXiv 2020. https://doi.org/10.48550/arXiv.2007.10261.

[51] Pappu A, Paige B. Making graph neural networks worth it for low-data molecular machine learning. Prepr ArXiv 2020. https://doi.org/10.48550/arXiv.2011.12203.

[52] Nguyen CQ, Kreatsoulas C, Branson KM. Meta-learning inicialitazions for low-resource drug discovery. Prepr ChemRxiv 2020:1–10. https://doi.org/10.26434/chemrxiv.11981622.v1.

[53] Jiang Z, Gao J, Lv X. MetaP: meta pattern learning for one-shot knowledge graph completion. In: Proc. 44th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. ACM; 2021. p. 2232–6. https://doi.org/10.1145/3404835.3463086.

[54] Guo Z, Zhang C, Yu W, Herr J, Wiest O, Jiang M, et al. Few-shot graph learning for molecular property prediction. In: Proc. Web Conf. 2021. ACM; 2021. p. 2559–67. https://doi.org/10.1145/3442381.3450112.

[55] Barrett R, White AD. Investigating active learning and meta-learning for iterative peptide design. J Chem Inf Model 2021;61:95–105. https://doi.org/10.1021/acs.jcim.0c00946.

[56] Wang J, Zheng S, Chen J, Yang Y. Meta learning for low-resource molecular optimization. J Chem Inf Model 2021;61:1627–36. https://doi.org/10.1021/acs.jcim.0c01416.

[57] Yao H., Wei Y., Huang L-K, Xue D., Huang J., Li Z. Functionally Regionalized Knowledge Transfer for Low-resource Drug Discovery. In: Marc'Aurelio Ranzato; Alina Beygelzimer; Yann Dauphin; Percy S. Liang; Jenn Wortman Vaughan, editor. Adv. Neural Inf. Process. Syst. 34 - 35th Conf. Neural Inf. Process. Syst. 10th ed., Neural information processing systems; 2021, p. 8256–68.

[58] Liu Y, Wu Y, Shen X, Xie L. COVID-19 multi-targeted drug repurposing using few-shot learning. Front Bioinforma 2021;1:1–10. https://doi.org/10.3389/fbinf.2021.693177.

[59] Ma J, Fong SH, Luo Y, Bakkenist CJ, Shen JP, Mourragui S, et al. Few-shot learning creates predictive models of drug response that translate from high-throughput screens to individual patients. Nat Cancer 2021;2:233–44. https://doi.org/10.1038/s43018-020-00169-2.

[60] Seidl P, Renz P, Dyubankova N, Neves P, Verhoeven J, Wegner JK, et al. Improving few- and zero-shot reaction template prediction using modern hopfield networks. J Chem Inf Model 2021. https://doi.org/10.1021/acs.jcim.1c01065.

[61] Wang Y, Abuduweili A, Yao Q, Dou D. Property-aware relation networks for few-shot molecular property prediction. Prepr Arxiv 2021:1–14. https://doi.org/10.48550/arXiv.2107.07994.

[62] Yao H, Wang Y, Wei Y, Zhao P, Mahdavi M, Lian D, et al. Meta-learning with an adaptive task scheduler. Prepr ArXiv 2021:1–13. https://doi.org/10.48550/arXiv.2110.14057.

[63] Gao P, Liu Z, Zhang J, Wang JA, Henkelman G. A fast, low-cost and simple method for predicting atomic/inter-atomic properties by combining a low dimensional deep learning model with a fragment based graph convolutional network. Crystals 2022;12:1740. https://doi.org/10.3390/cryst12121740.

[64] Gull S, Minhas F. AMP0: species-specific prediction of anti-microbial peptides using zero and few shot learning. IEEE/ACM Trans Comput Biol Bioinforma 2022;19:275–83. https://doi.org/10.1109/TCBB.2020.2999399.

[65] Yao S, Feng Z, Song J, Jia L, Zhong Z, Song M. Chemical property relation guided few-shot molecular property prediction. In: Proc. 2022 Int. Jt. Conf. Neural Networks. IEEE; 2022. p. 1–8. https://doi.org/10.1109/IJCNN55064.2022.9892419.

[66] Deng Y, Qiu Y, Xu X, Liu S, Zhang Z, Zhu S, et al. META-DDIE: predicting drug–drug interaction events with few-shot learning. Brief Bioinform 2022;23:1–8. https://doi.org/10.1093/bib/bbab514.

[67] He W, Jiang Y, Jin J, Li Z, Zhao J, Manavalan B, et al. Accelerating bioactive peptide discovery via mutual information-based meta-learning. Brief Bioinform 2022;23. https://doi.org/10.1093/bib/bbab499.

[68] Wang Z, Combs SA, Brand R, Calvo MR, Xu P, Price G, et al. LM-GVP: an extensible sequence and structure informed deep learning framework for protein property prediction. Sci Rep 2022;12:6832. https://doi.org/10.1038/s41598-022-10775-y.

[69] Mi X, Shukla D. Predicting the activities of drug excipients on biological targets using one-shot learning. J Phys Chem B 2022;126:1492–503. https://doi.org/10.1021/acs.jpcb.1c10574.

[70] Liu S, Qu M, Zhang Z, Cai H, Tang J. Structured multi-task learning for molecular property prediction. Prepr Arxiv 2022. https://doi.org/10.48550/arXiv.2203.04695.

[71] Sanchez-Fernandez A, Rumetshofer E, Hochreiter S, Klambauer G. CLOOME: contrastive learning unlocks bioimaging databases for queries with chemical structures. Prepr BioRxiv 2022:0–17. https://doi.org/10.1101/2022.11.17.516915.

[72] Schimunek J, Seidl P, Friedrich L, Kuhn D, Rippmann F, Hochreiter S, et al. Context-enriched molecule representations improve few-shot drug discovery. Prepr ArXiv 2023. https://doi.org/10.48550/arXiv.2305.09481.

[73] Lv Q, Zhou J, Yang Z, He H, Chen CYC. 3D graph neural network with few-shot learning for predicting drug-drug interactions in scaffold-based cold start scenario. Neural Networks 2023. https://doi.org/10.1016/j.neunet.2023.05.039.

[74] Li T, Shetty S, Kamath A, Jaiswal A, Jiang X, Ding Y, et al. CancerGPT: few-shot drug pair synergy prediction using large pre-trained language models. Prepr ArXiv 2023. https://doi.org/10.48550/arXiv.2304.10946.

[75] Seidl P, Vall A, Hochreiter S, Klambauer G. Enhancing activity prediction models in drug discovery with the ability to understand human language. Prepr ArXiv 2023. https://doi.org/10.48550/arXiv.2303.03363.

[76] Torres L, Arrais JP, Ribeiro B. Few-shot learning via graph embeddings with convolutional networks for low-data molecular property prediction. Neural Comput Appl 2023;35:13167–85. https://doi.org/10.1007/s00521-023-08403-5.

[77] Torres LHM, Ribeiro B, Arrais JP. Few-shot learning with transformers via graph embeddings for molecular property prediction. Expert Syst Appl 2023;225:120005. https://doi.org/10.1016/j.eswa.2023.120005.

[78] Ju W, Liu Z, Qin Y, Feng B, Wang C, Guo Z, et al. Few-shot molecular property prediction via hierarchically structured learning on relation graphs. Neural Networks 2023;163:122–31. https://doi.org/10.1016/j.neunet.2023.03.034.

[79] Lv Q, Chen G, Yang Z, Zhong W, Meta Chen CY-C. Learning with graph attention networks for low-data drug discovery. IEEE Trans Neural Networks Learn Syst 2023:1–13. https://doi.org/10.1109/TNNLS.2023.3250324.

[80] Lu Q, Zhang R, Zhou H, Ni D, Xiao W, Li J. MetaHMEI: meta-learning for prediction of few-shot histone modifying enzyme inhibitors. Brief Bioinform 2023;24. https://doi.org/10.1093/bib/bbad115.

[81] Meng Z, Li Y, Zhao P, Yu Y, King I. Meta-learning with motif-based task augmentation for few-shot molecular property prediction. In: Proc. 2023 SIAM Int. Conf. Data Min. Society for Industrial and Applied Mathematics; 2023. p. 811–9. https://doi.org/10.1137/1.9781611977653.ch91.
[82] Dong L (Leon), Qian Y, Gonzalez P, Öz OK, Sun X. Advancing drug discovery with deep learning: harnessing reinforcement learning and one-shot learning for molecular design in low-data situations. ACM SIGAPP Appl Comput Rev 2023;23:36–48. https://doi.org/10.1145/3594264.3594267.
[83] Vella D, Ebejer JP. Few-shot learning for low-data drug discovery. J Chem Inf Model 2023;63:27–42. https://doi.org/10.1021/acs.jcim.2c00779.
[84] Zhang Q, Zhang S, Feng Y, Shi J. Few-shot drug synergy prediction with a prior-guided hypernetwork architecture. IEEE Trans Pattern Anal Mach Intell 2023:1–17. https://doi.org/10.1109/TPAMI.2023.3248041.
[85] Chen K, Chen G, Li J, Huang Y, Wang E, Hou T, et al. MetaRF: attention-based random forest for reaction yield prediction with a few trails. J Cheminform 2023;15:43. https://doi.org/10.1186/s13321-023-00715-x.
[86] Rodríguez-Pérez R, Bajorath J. Multitask machine learning for classifying highly and weakly potent kinase inhibitors. ACS Omega 2019;4:4367–75. https://doi.org/10.1021/acsomega.9b00298.
[87] Bateni P, Barber J, Goyal R, Masrani V, van de Meent JW, Sigal L, et al. Beyond simple meta-learning: multi-purpose models for multi-domain, active and continual few-shot learning. Prepr Arxiv 2022. https://doi.org/10.48550/arXiv.2201.05151.
[88] Tan M. Prediction of anti-cancer drug response by kernelized multi-task learning. Artif Intell Med 2016;73:70–7. https://doi.org/10.1016/j.artmed.2016.09.004.
[89] Wenzel J, Matter H, Schmidt F. Predictive multitask deep neural network models for ADME-Tox properties: learning from large data sets. J Chem Inf Model 2019;59:1253–68. https://doi.org/10.1021/acs.jcim.8b00785.
[90] Li X, Xu Y, Lai L, Pei J. Prediction of human cytochrome P450 inhibition using a multitask deep autoencoder neural network. Mol Pharm 2018;15:4336–45. https://doi.org/10.1021/acs.molpharmaceut.8b00110.
[91] Taylor CJ, Felton KC, Wigh D, Jeraal MI, Grainger R, Chessari G, et al. Accelerated chemical reaction optimization using multi-task learning. ACS Cent Sci 2023;9:957–68. https://doi.org/10.1021/acscentsci.3c00050.
[92] Lin S, Shi C, Chen J. GeneralizedDTA: combining pre-training and multi-task learning to predict drug-target binding affinity for unknown drug discovery. BMC Bioinformatics 2022;23:367. https://doi.org/10.1186/s12859-022-04905-6.
[93] Wang X, Zhu H, Chen D, Yu Y, Liu Q, Liu Q. A complete graph-based approach with multi-task learning for predicting synergistic drug combinations. Bioinformatics 2023;39. https://doi.org/10.1093/bioinformatics/btad351.
[94] Rosenbaum L, Dörr A, Bauer MR, Boeckler FM, Zell A. Inferring multi-target QSAR models with taxonomy-based multi-task learning. J Cheminform 2013;5:33. https://doi.org/10.1186/1758-2946-5-33.
[95] Sharma A, Rani R. Drug sensitivity prediction framework using ensemble and multi-task learning. Int J Mach Learn Cybern 2020;11:1231–40. https://doi.org/10.1007/s13042-019-01034-0.
[96] Moon C, Kim D. Prediction of drug–target interactions through multi-task learning. Sci Rep 2022;12:18323. https://doi.org/10.1038/s41598-022-23203-y.
[97] Zhao Z, Qin J, Gou Z, Zhang Y, Yang Y. Multi-task learning models for predicting active compounds. J Biomed Inform 2020;108:103484. https://doi.org/10.1016/j.jbi.2020.103484.
[98] Sadawi N, Olier I, Vanschoren J, van Rijn JN, Besnard J, Bickerton R, et al. Multi-task learning with a natural metric for quantitative structure activity relationship learning. J Cheminform 2019;11:68. https://doi.org/10.1186/s13321-019-0392-1.
[99] Sung F, Yang Y, Zhang L, Xiang T, Torr PHS, Hospedales TM. Learning to compare: relation network for few-shot learning. In: Proc. 2018 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. IEEE; 2018. p. 1199–208. https://doi.org/10.1109/CVPR.2018.00131.
[100] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. Prepr ArXiv 2017. https://doi.org/10.48550/arXiv.1703.03400.
[101] Zhang T., Yu B. Boosting with early stopping: convergence and consistency 2005. https://doi.org/10.1214/009053605000000255.
[102] Nichol A, Achiam J, Schulman J. On first-order meta-learning algorithms. Prepr ArXiv 2018. https://doi.org/10.48550/arXiv.1803.02999.
[103] Schimunek J, Friedrich L, Kuhn D, Rippmann F, Hochreiter S, Klambauer G. A generalized framework for embedding-based few-shot learning methods in drug discovery. ELLIS Mach Learn Mol Discov Work 2021;21:1–13.
[104] Bargagli Stoffi FJ, Cevolani G, Gnecco G. Simple models in complex worlds: Occam's razor and statistical learning theory. Minds Mach 2022;32:13–42. https://doi.org/10.1007/s11023-022-09592-z.
[105] Gamberger D., Lavrač N. Conditions for Occam's razor applicability and noise elimination, 1997, p. 108–23. https://doi.org/10.1007/3-540-62858-4_76.