

[Artificial Intelligence in the Life Sciences 4 (2023) 10008](https://doi.org/10.1016/j.ailsci.2023.100086)6

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  | Contents lists available at ScienceDirect |  |
| Artificial Intelligence in the Life Sciences |
| journal homepage: [www.elsevier.com/locate/ailsci](https://www.elsevier.com/locate/ailsci) |

Review

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

Sabrina Silva-Mendonçaa,b, Arthur Ricardo de Sousa Vit´oriab,c, Telma Woerle de Limab,c, Arlindo Rodrigues Galv˜ao-Filhob,c, Carolina Horta Andradea,b,\*

a *Laboratory for Molecular Modeling and Drug Design (LabMol), Faculty of Pharmacy, Universidade Federal de Goi*´*as, Goi*ˆ*ania, GO 74605-170, Brazil* b *Center for Excellence in Artificial Intelligence (CEIA), Institute of Informatics, Universidade Federal de Goi*´*as, Goi*ˆ*ania, GO 74605-170, Brazil* c *Institute of Informatics, Universidade Federal de Goi*´*as, Goi*ˆ*ania, Brazil*

|  |  |
| --- | --- |
| A R T I C L E I N F O | A B S T R A C T |
| *Keywords:*  Drug discovery  Low data  Few-shot learning  Meta-learning  Neglected diseases  Rare diseases | 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 artifi-cial 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, re-searchers can efficiently identify potential drug targets and develop new drug candidates. Among the AI tech-niques, 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 ma-chines [4]. Essentially, AI encompasses the capacity of machines to carry

\* Corresponding author.

*E-mail address:* [carolina@ufg.br](mailto:carolina@ufg.br) (C.H. Andrade).

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 develop-ment 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

<https://doi.org/10.1016/j.ailsci.2023.100086>   
[Received 30 June 2023; Received in revised f](https://doi.org/10.1016/j.ailsci.2023.100086)orm 29 August 2023; Accepted 4 September 2023   
Available online 9 September 2023   
[2667-3185](http://creativecommons.org/licenses/by-nc-nd/4.0/)/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

*S. Silva-Mendonça et al.*  *Artificial Intelligence in the Life Sciences 4 (2023) 100086*

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 uti-lized in scenarios where data resources are limited. It aims to achieve satisfactory learning performance despite the constraints of limited su-pervised 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 per-formance of models in FSL scenarios.

The Transfer Learning (TL) involves leveraging pre-existing knowl-edge 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 net-works, 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.

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 ap-proaches [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, architec-tures, type of prediction, taxonomy, metrics, and year of publication.

**Few-Shot learning**  **Recent advances in FSL for drug discovery**

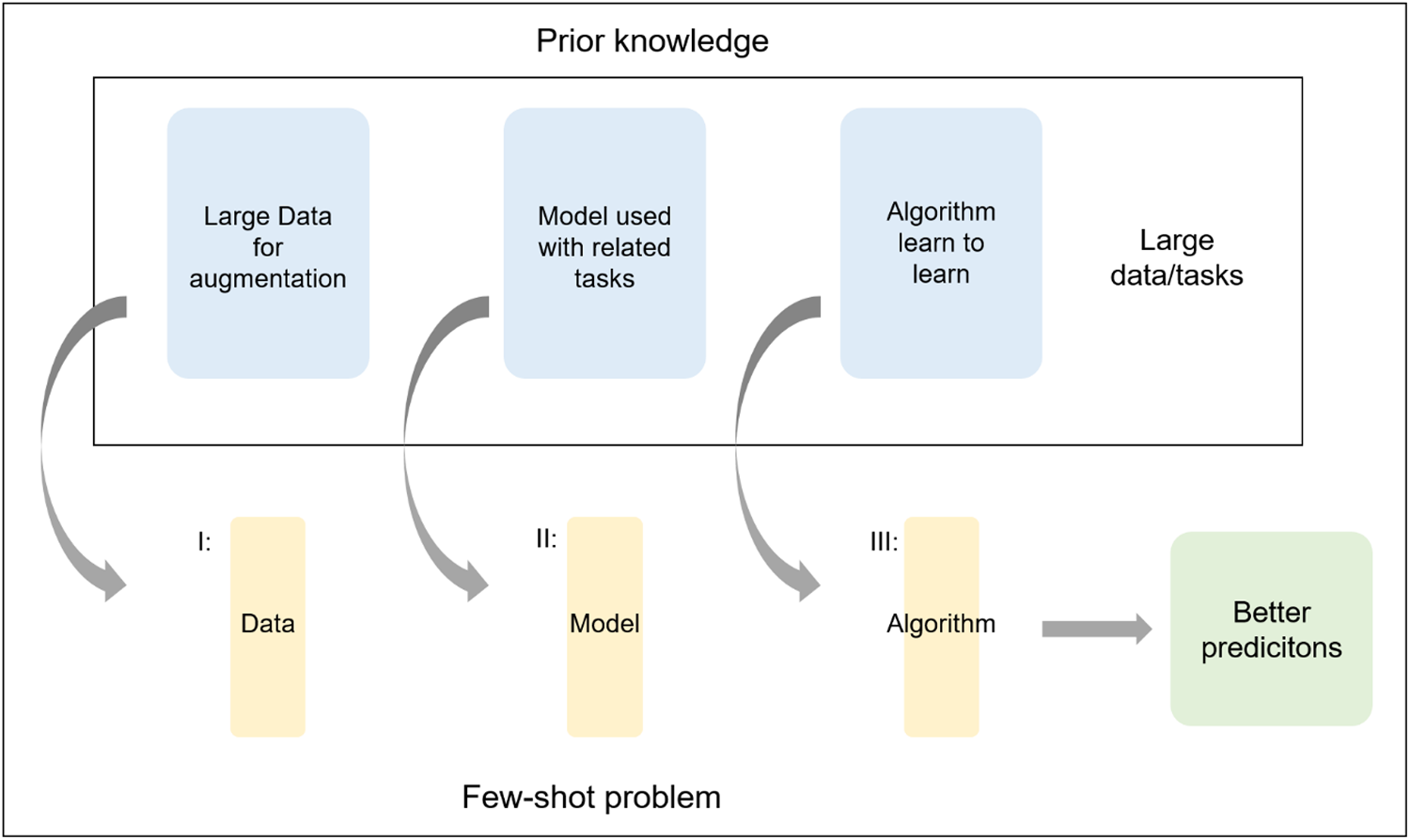
Few-Shot Learning (FSL) can be employed in various types of data for supervised learning problems. These include: (*i*) classification: this type of FSL approach, known as *N*-way-*K*-shot, involves considering *N* from the number of classes, and *K* is the number of examples. Alternatively, one-shot learning (OSL) is employed when only one example is avail-able, 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

2

*S. Silva-Mendonça et al.*  *Artificial Intelligence in the Life Sciences 4 (2023) 100086*



**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 vocab-ulary. Using convex combinations, the approach creates prototype clusters to represent each class. Then, the authors evaluated the effec-tiveness 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 environ-ments [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 ach-ieved 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 genera-tive models. The initial classification is multi-task learning, which in-volves 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.

3

*S. Silva-Mendonça et al.*  *Artificial Intelligence in the Life Sciences 4 (2023) 100086*

**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) Hit top 1% (21.52) | 2017 |
| [49] | ENZYME database, Kinase. com, KEGG database | ProtVec representation | LSTM and Recurrent Neural Network | Kinase-  phosphosite associations Biological  activity | Model | 2020 |
| [50] | DrugBank, GNBR, Hetionet, STRING, IntAct, and DGIdb databases  ChEMBL20 database | Text embeddings | Relacional GCN | Model | Hit top 1% (25.32) | 2020 |
| [51] | 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)  AUC-PR  CHEMBL ID 918,058  (0.747 ± 0.076)  CHEMBL ID 1,804,798  (0.369 ± 0.021)  Hit top 1% (0.188) | 2020 |
| [52] | ChEMBL20 database | Molecular graphs | Gated GCN and Model  Agnostic Meta-Learning | Biological activity | Algorithm | 2020 |
| [53] | NELL-One, COVID19-One | KGC embeddings | Meta pattern learning framework | Built KGC  dataset  COVID19-One Molecular  property | Data | 2021 |
| [54] | Tox21 and SIDER | Molecular Graphs and SMILES | Model agnostic machine learning and GCN | Algorithm | AUC-ROC  Tox21 1-shot (0.77)  SIDER 1-shot (0.74)  Tox21 5-shot (0.78)  SIDER 5-shot (0.75)  AUC-PR  All enzymes (0.206 ± 0.008) ACC  The best: anti-MRSA (0.93 ±0.02)  The worst: soluble (0.50  ±0.02)  Averaged success rate  (49.39%)  Averaged accuracy SIDER  (70.01 ± 0.86%)  Tox21 (71.07 ± 0.91%)  MUV (60.66 ± 1.09%)  ToxCast (74.02 ± 1.57%)  AUC-ROC  Jak1 8:1:1 ratio (0.99 ± 0.11) F1  Jak1 8:1:1 ratio (0.92 ± 0.80) KU-55,933 resistance (third from top; *R* = 0.54) | 2021 |
| [28] | ChEMBL27 database | Molecular Graphs | Model Agnostic Meta- learning  GCN | Biological  activity  Peptide design | Model | 2021 |
| [55] | Antimicrobial Peptide  Database (APD), PROSO II database, TULA-2 and SHP-2 protein library. | Peptide sequence | Algorithm | 2021 |
| [56] | ChEMBL20 database | Molecular Graphs | Model Agnostic Meta- learning  Model Agnostic Meta- learning and recurrent neural network | Biological activity  Molecular property | Algorithm | 2021 |
| [57] | ChEMBL database | RDKit descriptors | Algorithm | 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 | 2021 |
| [59] | Cancer Cell Line  Encyclopedia (CCLE) project from the DepMap website United States Patent and  Trademark Office (USPTO) dataset | Molecular Graphs | Multi-layer perceptron and Model Agnostic Meta-  learning  Modern Holpfield Network | Biological activity | Algorithm | 2021 |
| [60] | Weighted ECFP | Reaction  prediction | Algorithm | Top-k Accuracy (%) USPTO- lg  TOP 1 (16.9)  TOP 10 (42.2)  TOP 100 (72.4)  AUC-ROC (%) 1-shot  Tox21 (83.01 ± 0.09)  SIDER (74.46 ± 0.29)  MUV (66.94 ± 1.12)  ToxCast (73.63 ± 1.00)  Accuracy ± 95% confidence interval (0.252) | 2021 |
| [61] | MoleculeNet | Molecular Graphs | Model Agnostic Meta- Learning and GCN | Molecular property | Model | 2021 |
| [62] | miniImagenet, Tox21, SIDER, MUV, ToxCast | Embeddings | Bi-directional LSTM,  Multi-layer perceptron, and GCN  Fragment-based GCN | Biological activity | Model | 2021 |
| [63] | Alchemy library | Multiple-level molecular fragments, RDKit  descriptors, and QM  calculated descriptors  Feature vector | 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  Tox21 and SIDER | Multi-Layer Perceptron | Biological activity | Algorithm | AUC-ROC  Between 0.6 to 0.96 | 2022 |
| [65] | 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*)

4

*S. Silva-Mendonça et al.*  *Artificial Intelligence in the Life Sciences 4 (2023) 100086*

**Table 1** (*continued*)

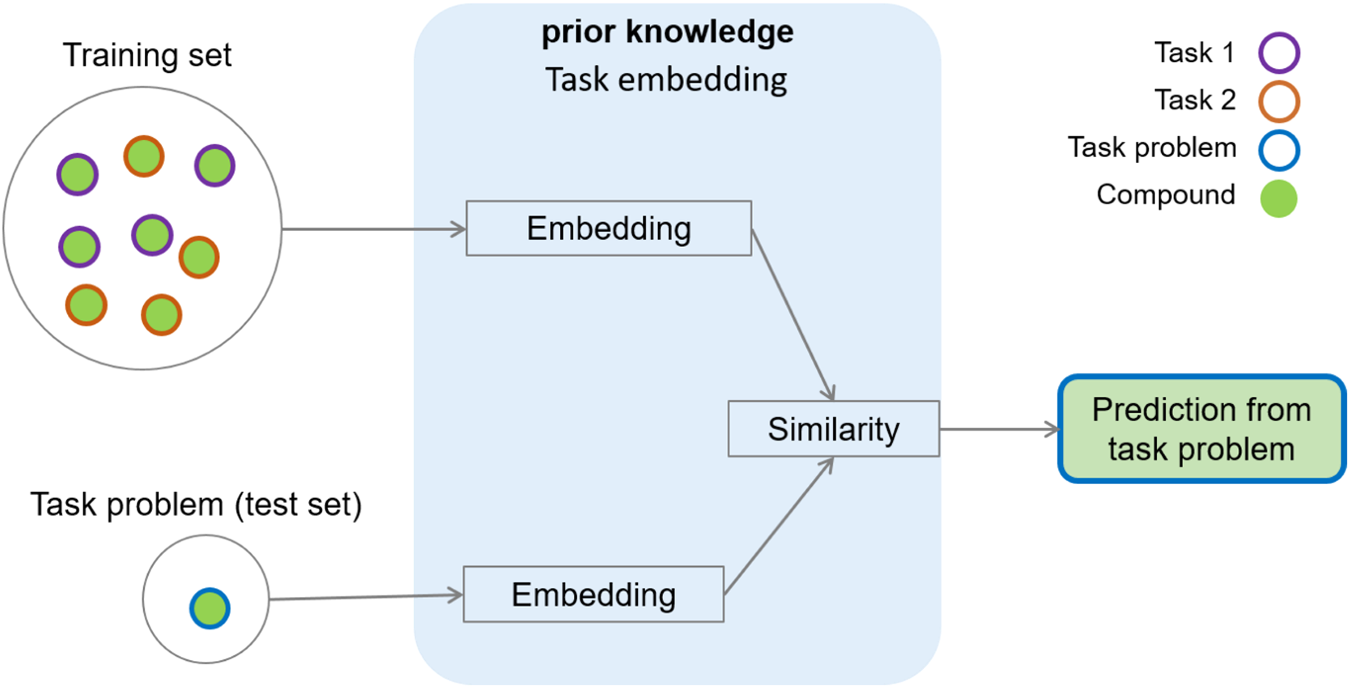
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Authors | Dataset | Features | Architecture | Prediction | Taxonomy | Metrics | Year | |
| [66] | Drug Bank | SMILES | GCN | Drug-drug  interaction | Model | 5-shot Tox21 (78.81)  5-shot SIDER (77.60)  Accuracy average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

5

*S. Silva-Mendonça et al.*  *Artificial Intelligence in the Life Sciences 4 (2023) 100086*

**Table 1** (*continued*)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Authors | Dataset | Features | Architecture | Prediction | Taxonomy | Metrics | Year |
| [84] | SYNERGxDB database | Molecular graphs | Hypernetwork, Model Agnostic Meta-learning, and GCN | Drug-drug  interaction | Model | matchingNet (0.352 ± 0.121) prototypicalNet (0.373 ± 0.102)  relationNet (0.342 ± 0.093) 5-shot  MSE (0.115 ± 0.002) SCC (0.508 ± 0.004) R2 (0.180 ± 0.008) Suzuki-Miyaura HTE  Dataset  RSME (18.3102)  MAE (0.6350)  R2 (13.8344) | 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 | 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 em-

beddings 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 ar-chitecture that effectively captures the structural and functional infor-mation of both drugs and proteins. The GCN model learns to encode the molecular structures and interactions into a low-dimensional embed-ding 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 contras-tive learning, a self-supervised learning approach, to establish associa-tions between chemical structures and corresponding bioimages. The neural network is trained to maximize similarity between paired in-stances 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 seman-tically linked. By integrating chemical structure information with bio-imaging data, CLOOME offers new opportunities for efficient and targeted analysis, allowing researchers to leverage large-scale bio-imaging databases for drug discovery and biological research [71].

A closely related approach to QSAR is MetaHMEI, described by Lu

6

*S. Silva-Mendonça et al.*  *Artificial Intelligence in the Life Sciences 4 (2023) 100086*

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 pub-lished 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 Hop-field Networks (MHopNets) to improve the prediction of reaction tem-plates 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 Met-aRF, 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 out-performs 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 pro-cesses and accelerated drug discovery efforts [85].

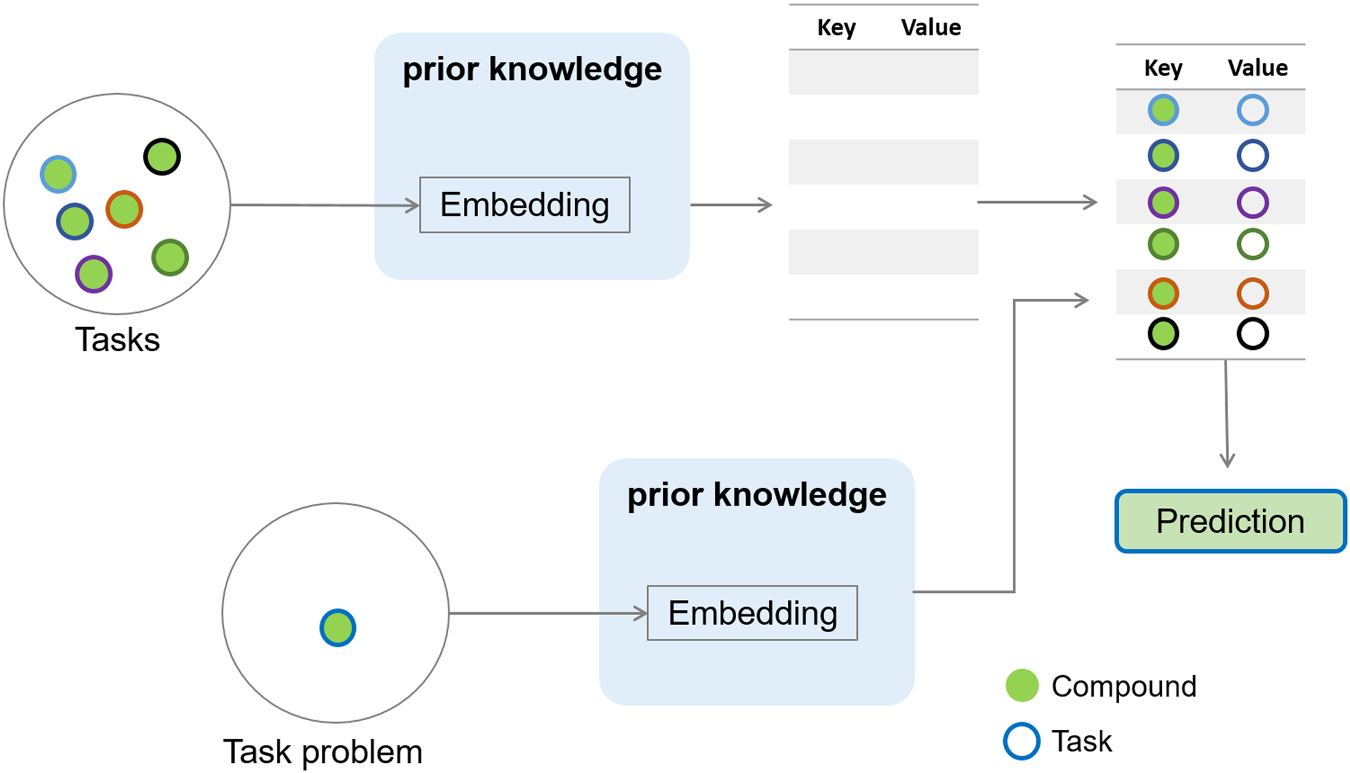
Jiang and Gao [53] presented the MetaP framework, which consists

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 em-ploys 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 out-performs existing state-of-the-art methods in terms of accuracy and ef-ficiency, 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 addi-tional samples that serve to augment the available data for the target task.

An example of generative modeling is shown by Dong and co- workers [82], who introduced a novel approach to address the chal-lenge 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 effec-tiveness 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.

7

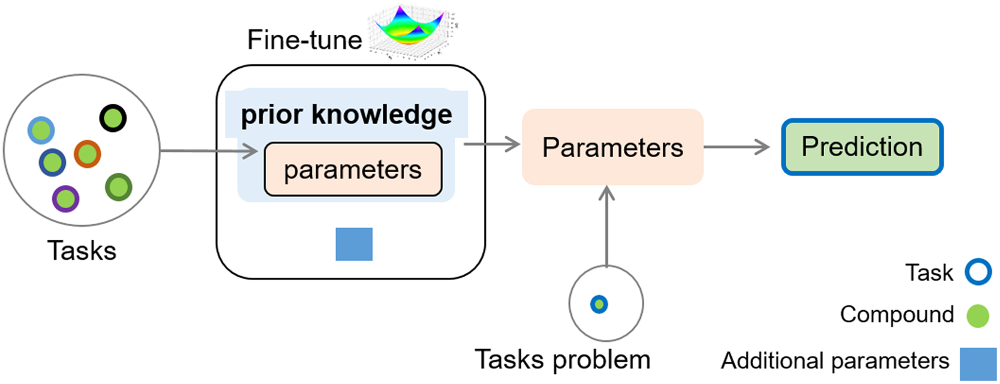
*S. Silva-Mendonça et al.*  *Artificial Intelligence in the Life Sciences 4 (2023) 100086*

(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 understand-ing. 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, empow-ering CancerGPT to generalize its understanding and make precise predictions for previously unseen drug combinations. Moreover, experimental evaluations on various cancer cell line datasets demon-strated that CancerGPT surpasses existing methods in predicting drug pair synergy with FSL [74]. The approach showcases remarkable accu-racy 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 dis-covery 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 itera-tively refining the existing parameters, the model can enhance its per-formance and exhibit improved generalization when faced with new classes or tasks that have limited labeled examples (Fig. 4). The refine-ment process commonly employs gradient-based optimization tech-niques, 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 regu-larization in FSL refers to the process of refining the pre-trained pa-rameters 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].

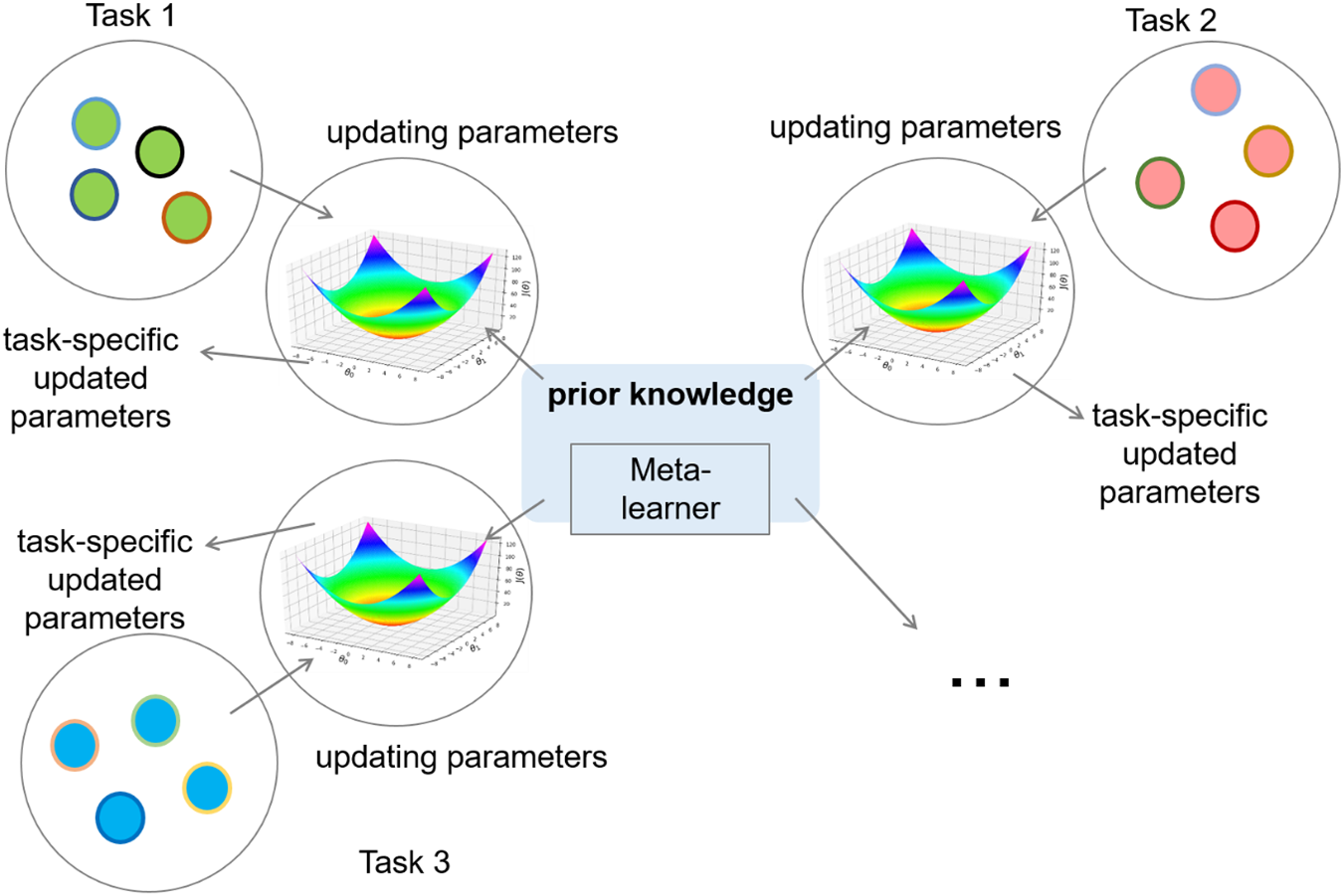
Gull and Minhas [64] proposed a technique for the accurate pre-diction of antimicrobial peptides (AMPs) in different species using zero and FSL. AMPs are important components of the innate immune system and exhibit potential therapeutic properties. However, the diversity of AMPs across various species 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



**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, ac-cording to the necessity of additional parameters from the training set.

8

*S. Silva-Mendonça et al.*  *Artificial Intelligence in the Life Sciences 4 (2023) 100086*



**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.

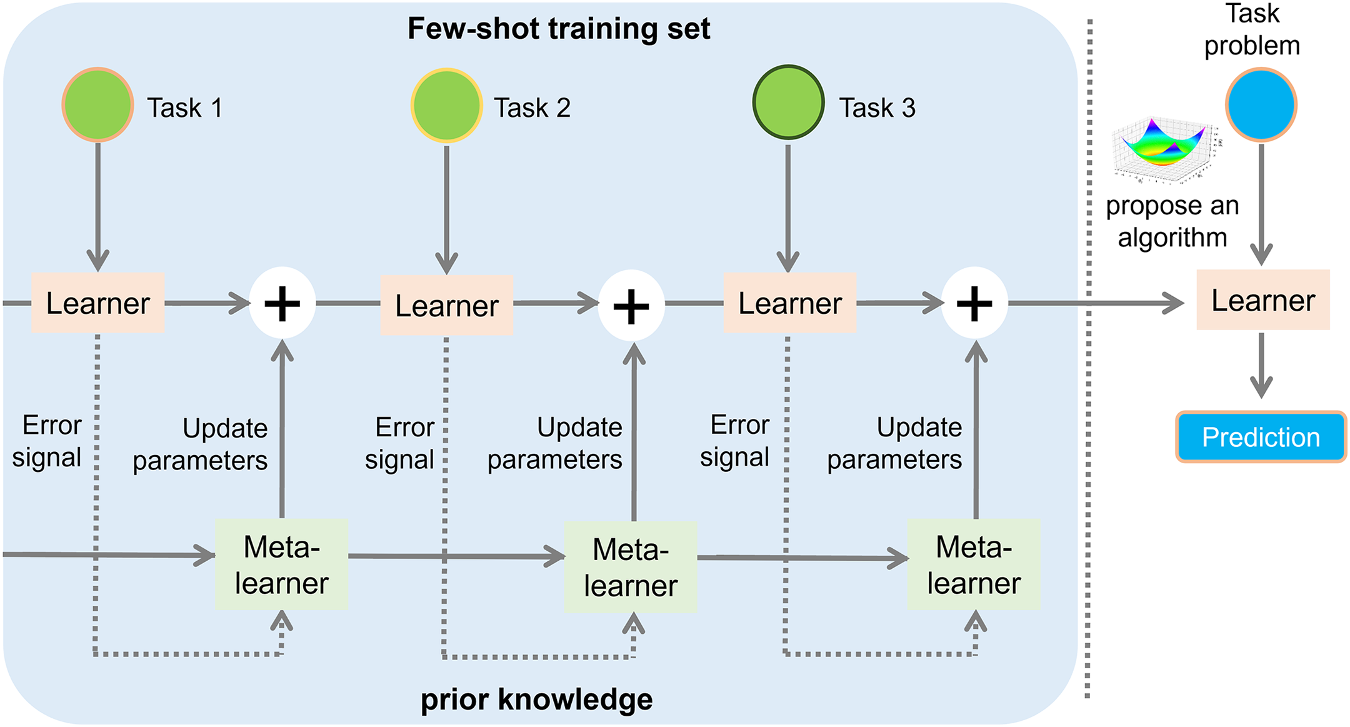
Pappu and Paige [51] suggested an additional noteworthy example that harnesses the power of GNNs to effectively model the intricate re-lationships and structures of molecules. However, the efficacy of GNNs is often hindered by the scarcity of labeled data. To overcome this limitation, this study proposed an innovative framework that enhances the performance of GNNs in low-data scenarios. The technique in-tegrates 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 po-tential 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 accu-rate 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 effec-tively 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

9

*S. Silva-Mendonça et al.*  *Artificial Intelligence in the Life Sciences 4 (2023) 100086*



**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 proper-ties, 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 tech-nique that utilizes mutual information, a statistical dependence mea-sure, to establish connections between peptide sequences and their bioactivity. This approach involves extracting informative features from the peptide sequences, including amino acid composition, physico-chemical properties, and structural motifs, to identify potential bioac-tive 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 augmen-tation 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 solu-tion 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 acceler-ating 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] intro-duced 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 effec-tively 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

10

*S. Silva-Mendonça et al.*  *Artificial Intelligence in the Life Sciences 4 (2023) 100086*

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 chal-lenges, 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 IC50 and/or EC50 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 poten-tial 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 pre-dictions. 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 fair-ness 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.

11

*S. Silva-Mendonça et al.*  *Artificial Intelligence in the Life Sciences 4 (2023) 100086*

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 tar-gets 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 effec-tive based on their chemical properties and other characteristics. 3. Designing new drugs: Another potential application of FSL in drug dis-covery 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 re-sults for FSL tasks. The appeal of this approach lies in its straightforward calculation of similarities between embeddings, as opposed to the uti-lization 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 re-sources. 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 prob-lems, assess drug synergy probabilities, and evaluate drug toxicity and ADME properties. Such advancements may ultimately lead to a reduc-tion 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.

12

*S. Silva-Mendonça et al.*  *Artificial Intelligence in the Life Sciences 4 (2023) 100086*

[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 Pat](https://doi.org/10.1093/bib/bbac532)´[on A, e](https://doi.org/10.1093/bib/bbac532)t 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., Bronsk[ill J.F., Maziarz K., Misztela H., Lanini J](https://doi.org/10.3390/ijms24021146)., 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, Wan[g D. A survey of transfer learning. J Big Data](https://doi.org/10.1007/s10994-017-5685-x) 2016;3:9. <https://doi.org/10.1186/s40537-016-0043-6>.

[31] Vilalta R, [Drissi Y. A perspective view and survey of m](https://doi.org/10.1186/s40537-016-0043-6)eta-learning. Artif Intell Rev 2002;18:77–95. <https://doi.org/10.1023/A:1019956318069>.

[32] Hospedales TM, Anto[niou A, Micaelli P, Storkey AJ. Meta-learni](https://doi.org/10.1023/A:1019956318069)ng in neural networks: a survey. IEEE Trans Pattern Anal Mach Intell 2021. [https://doi.org/](https://doi.org/10.1109/TPAMI.2021.3079209) [10.1109/TPAMI.2021.3079209](https://doi.org/10.1109/TPAMI.2021.3079209). 1–1.

[33] [Qiao S, Liu C, Shen W, Yuille A](https://doi.org/10.1109/TPAMI.2021.3079209). Few-shot image recognition by predicting parameters from activations. Prepr ArXiv 2017. [https://doi.org/10.48550/](https://doi.org/10.48550/arXiv.1706.03466) [arXiv.1706.03466](https://doi.org/10.48550/arXiv.1706.03466).

[34] [Larochelle H, Erh](https://doi.org/10.48550/arXiv.1706.03466)[an D, Bengio Y. Zero-data learning of new tasks. In: Proc AAAI](http://refhub.elsevier.com/S2667-3185(23)00030-2/sbref0034) [2008 Proc Twenty-Third AAAI Conf Artif Intell. 2; 2008. p. 646–51.](http://refhub.elsevier.com/S2667-3185(23)00030-2/sbref0034)

[35] [Zell A., Sumbul G., Demir B. Deep Metric Learning-Based Semi-Sup](http://refhub.elsevier.com/S2667-3185(23)00030-2/sbref0034)ervised 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](http://refhub.elsevier.com/S2667-3185(23)00030-2/sbref0036) [visual concepts. In: Proc 33rd Annu Meet Cogn Sci Soc CogSci 2011. 33; 2011.](http://refhub.elsevier.com/S2667-3185(23)00030-2/sbref0036) [p. 2568–73.](http://refhub.elsevier.com/S2667-3185(23)00030-2/sbref0036)

[37] [Wang Q, C](http://refhub.elsevier.com/S2667-3185(23)00030-2/sbref0036)hen 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.](https://doi.org/10.1016/j.neunet.2019.09.029) [neunet.2019.09.029](https://doi.org/10.1016/j.neunet.2019.09.029).

[38] [Lake BM, Salakhutdi](https://doi.org/10.1016/j.neunet.2019.09.029)nov R, Tnenbaum JB. Human-level concept learning through probabilistic program induction. Science (80-) 2015;350:1332–8. [https://doi.](https://doi.org/10.1126/science.aab3050)

[org/10.1126/science.aab3050](https://doi.org/10.1126/science.aab3050).

[39] [Fei-Fei Li, Fergus R, Perona P.](https://doi.org/10.1126/science.aab3050) One-shot learning of object categories. IEEE Trans Pattern Anal Mach Intell 2006;28:594–611. [https://doi.org/10.1109/](https://doi.org/10.1109/TPAMI.2006.79)   
[TPAMI.2006.79](https://doi.org/10.1109/TPAMI.2006.79).

[40] [Yin W. Meta-lea](https://doi.org/10.1109/TPAMI.2006.79)rning 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/](https://doi.org/10.1080/17460441.2022.2114451)   
[17460441.2022.2114451](https://doi.org/10.1080/17460441.2022.2114451).

[43] [Koch G, Zemel R, Salakh](https://doi.org/10.1080/17460441.2022.2114451)[utdinov R. Siamese neural networks for one-shot image](http://refhub.elsevier.com/S2667-3185(23)00030-2/sbref0043) [recognition. ICML Deep Learn. Work. 2015.](http://refhub.elsevier.com/S2667-3185(23)00030-2/sbref0043)

[44] [Snell J, Swersky K, Zemel R. Prototypical n](http://refhub.elsevier.com/S2667-3185(23)00030-2/sbref0043)etworks for few-shot learning. Prepr ArXiv 2017:4078–88. <https://doi.org/10.48550/arXiv.1703.05175>. 2017- Decem.

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

[46] [Zhao J, Lin X, Zhou J, Yang J, He L,](https://doi.org/10.48550/arXiv.2205.15745) 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](https://doi.org/10.1109/ICME46284.2020.9102809) adversarial networks. Prepr Arxiv 2020. [https://doi.org/10.48550/](https://doi.org/10.48550/arXiv.2010.11943) [arXiv.2010.11943](https://doi.org/10.48550/arXiv.2010.11943).

[48] [Altae-Tran H, Ram](https://doi.org/10.48550/arXiv.2010.11943)sundar 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/](https://doi.org/10.1021/acscentsci.6b00367) [acscentsci.6b00367](https://doi.org/10.1021/acscentsci.6b00367).

[49] [Deznabi I, Arabaci](https://doi.org/10.1021/acscentsci.6b00367) 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/](https://doi.org/10.1093/bioinformatics/btaa013) [btaa013](https://doi.org/10.1093/bioinformatics/btaa013).

[50] [Ioannidi](https://doi.org/10.1093/bioinformatics/btaa013)s VN, Zheng D, Karypis G. Few-shot link prediction via graph neural networks for Covid-19 drug-repurposing. Prepr ArXiv 2020. [https://doi.org/](https://doi.org/10.48550/arXiv.2007.10261) [10.48550/arXiv.2007.10261](https://doi.org/10.48550/arXiv.2007.10261).

[51] [Pappu A, Paige B. Making gr](https://doi.org/10.48550/arXiv.2007.10261)aph neural networks worth it for low-data molecular machine learning. Prepr ArXiv 2020. [https://doi.org/10.48550/](https://doi.org/10.48550/arXiv.2011.12203)   
[arXiv.2011.12203](https://doi.org/10.48550/arXiv.2011.12203).

[52] [Nguyen CQ, Kreat](https://doi.org/10.48550/arXiv.2011.12203)soulas C, Branson KM. Meta-learning inicialitazions for low- resource drug discovery. Prepr ChemRxiv 2020:1–10. [https://doi.org/10.26434/](https://doi.org/10.26434/chemrxiv.11981622.v1) [chemrxiv.11981622.v1](https://doi.org/10.26434/chemrxiv.11981622.v1).

[53] [Jiang Z, Gao J, Lv X. M](https://doi.org/10.26434/chemrxiv.11981622.v1)etaP: 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, Zha[ng C, Yu W, Herr J, Wiest O, Jiang M, et al](https://doi.org/10.1145/3404835.3463086). 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>.

13

*S. Silva-Mendonça et al.*  *Artificial Intelligence in the Life Sciences 4 (2023) 100086*

[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 (Le[on), Qian Y, Gonzalez P, ¨Oz OK, Sun X. Advancin](https://doi.org/10.1137/1.9781611977653.ch91)g 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](https://doi.org/10.1145/3594264.3594267) 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 predictio](https://doi.org/10.1021/acs.jcim.2c00779)n 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](https://doi.org/10.1109/TPAMI.2023.3248041). 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] Rodríg[uez-P](https://doi.org/10.1186/s13321-023-00715-x)´[erez R, Bajorath J. Multitask machine lea](https://doi.org/10.1186/s13321-023-00715-x)rning for classifying highly and weakly potent kinase inhibitors. ACS Omega 2019;4:4367–75. [https://doi.](https://doi.org/10.1021/acsomega.9b00298) [org/10.1021/acsomega.9b00298](https://doi.org/10.1021/acsomega.9b00298).

[87] [Bateni P, Barber J, Goyal R, Mas](https://doi.org/10.1021/acsomega.9b00298)rani 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/](https://doi.org/10.48550/arXiv.2201.05151) [arXiv.2201.05151](https://doi.org/10.48550/arXiv.2201.05151).

[88] [Tan M. Prediction](https://doi.org/10.48550/arXiv.2201.05151) 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 mode](https://doi.org/10.1016/j.artmed.2016.09.004)ls 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, La[i L, Pei J. Prediction of human cytochrome](https://doi.org/10.1021/acs.jcim.8b00785) 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,](https://doi.org/10.1021/acs.molpharmaceut.8b00110) 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 a](https://doi.org/10.1021/acscentsci.3c00050)nd 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>.

14