JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Open Access

pubs.acs.org/jcim

Application Note

MHNfs: Prompting In-Context Bioactivity Predictions for Low-Data Drug Discovery

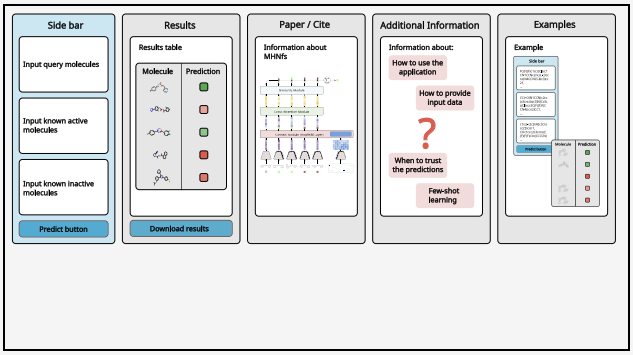
Johannes Schimunek, Sohvi Luukkonen, and Günter Klambauer*

Cite This: J. Chem. Inf. Model. 2025, 65, 4243−4250

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Today's drug discovery increasingly relies on computational and machine learning approaches to identify novel candidates, yet data scarcity remains a significant challenge. To address this limitation, we present *MHNfs*, an application specifically designed to predict molecular activity in low-data scenarios. At its core, *MHNfs* leverages a state-of-the-art few-shot activity prediction model, named MHNfs, which has demonstrated strong performance across a large set of prediction tasks in the benchmark data set FS-Mol. The application features an intuitive interface that enables users to prompt the model for precise activity predictions based on a small number of known active and inactive molecules, akin to interactive interfaces for large language models. To evaluate its efficacy, we simulate real-world scenarios by recasting PubChem bioassays as few-shot prediction tasks. *MHNfs* offers a streamlined and accessible solution for deploying advanced few-shot learning models, providing a valuable tool for accelerating drug discovery.



INTRODUCTION

Bioactivity Prediction Models Are Indispensable Tools in Modern Drug Discovery Projects.[2−5] The process of identifying new drug candidates is fraught with challenges: it is inherently complex, expensive, time-consuming, and prone to failure.[3,6] These difficulties arise from the intricate nature of biological systems,[7] the vast and largely unexplored chemical space of potential drug candidates,[8−11] and the high costs associated with experimental validation and clinical trials.[3,12] To improve the efficiency and cost-effectiveness of this process, machine learning (ML) methods have been increasingly adopted across various stages of the drug discovery pipeline,[7,13] demonstrating the ability to reduce both time and monetary investment.[14,15] Among these methods, bioactivity prediction models, which map molecular structure to bioactivity using known data, have emerged as key enablers.[16] These models are integral to large-scale virtual screening efforts,[17,18] enabling researchers to efficiently evaluate vast numbers of molecules. Today, a wide variety of such models exist, differing in architecture, training schemes, implementation, and application scope (see Table 1, Model Card1).

The Landscape of ML-Based Activity Prediction Models and Available Software. Machine learning models for activity prediction come in various forms, including descriptor-based multilayer perceptrons,[17,19] Random Forests (RFs),[20−23] gradient-boosting methods,[24,25] as well as graph neural networks.[26] To support researchers, several open source software packages[27−31] provide tools that cover the entire bioactivity prediction pipeline, from molecular data preprocessing to model development and evaluation. Although ML- and deep learning-based bioactivity models and their software are widely used in drug discovery, they are typically tailored to specific tasks that were seen during training. Adapting them to new tasks often requires retraining, which usually demands substantial data, limiting their use in data-scarce scenarios.

Few-Shot Learning Methods for Low-Data Drug Discovery. Drug discovery projects often start with only a few known molecules,[18,32,33] whereas standard bioactivity prediction models require large data sets for training.[17,19,34−36] To address this gap, few-shot learning methods have been developed for low-data scenarios.[37−41] Among them, `MHNfs`[42] represents the state-of-the-art that has been trained and evaluated on the FS-Mol benchmark data set.[32] When provided with a so-called support set of a few active and inactive molecules, `MHNfs` predicts the molecular activity of query molecules by comparing learned molecule representations and weighting the support set's labels based on similarity.

Received: December 19, 2024
Revised: April 9, 2025
Accepted: April 10, 2025
Published: April 30, 2025

Table 1. Model Card[1] - *MHNfs* on HuggingFace

Model Details

- Developed by researchers at the Johannes Kepler University Linz, 2023, v1
- Embedding-based few-shot method based on Modern Hopfield Networks (Transformer architecture)
- Deployed on HuggingFace: https://huggingface.co/spaces/ml-jku/mhnfs

Intended Use

- Intended to be used to predict bioactivities for targets similar to the targets included in the FS-Mol benchmark
- User input: Known actives and inactives, and molecules to be predicted (SMILES)
- Model output: Bioactivity table for requested molecules

Factors

- The performance of the model depends on the target family.
- Evaluated target families: Oxidoreductases, kinases, hydrolases, lysases, isomerases, ligases, translocases

Metrics, Training and Evaluation Data

- Training and evaluation data: FS-Mol
- External evaluation data: PubChem assays
- Reported metrics:
  - FS-Mol: AUC, ΔAUC-PR
  - PubChem: AUC, ΔAUC-PR, MCC, BACC, and BEDROC

Caveats and Recommendations

- Use model for oxidoreductases, kinases, hydrolases, lysases, isomerases, ligases, and translocases
- For other targets, check whether similar targets are included in the FS-Mol training data: https://github.com/microsoft/FS-Mol/blob/main/datasets/targets/target_info.csv

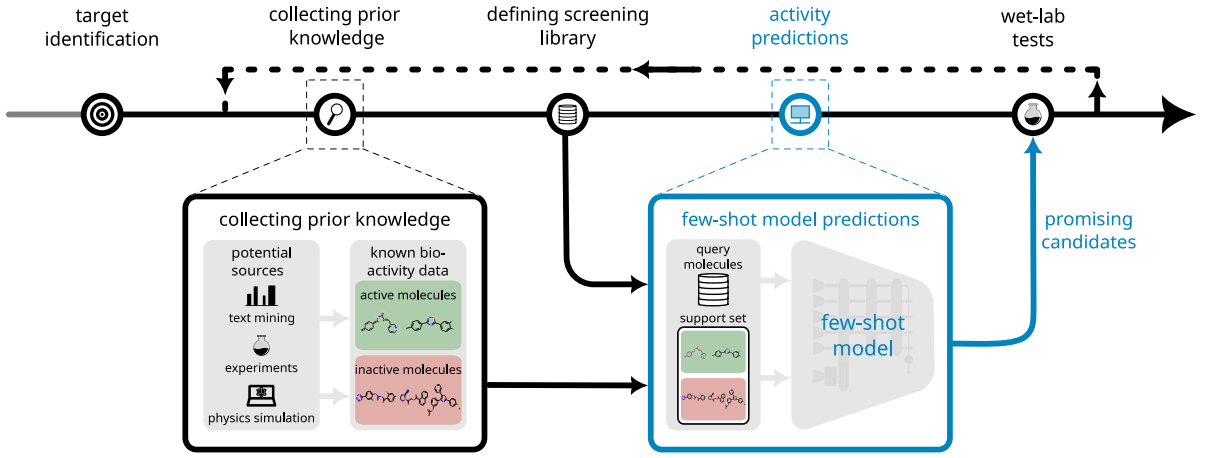


Figure 1. Early stage drug discovery pipeline and the role of *MHNfs*. *MHNfs* is a prompting-based few-shot bioactivity prediction application designed for very early stage virtual compound screening when prior knowledge about bioactivity is minimal.

Training few-shot learning models is resource-intensive and requires expert knowledge, and their use typically demands programming skills and familiarity with deep learning frameworks, limiting their adoption in chemoinformatics and drug discovery projects. Despite recent advancements in accuracy[42] and practical utility,[43] these models remain largely inaccessible to many researchers. However, like large language models (LLMs), they could become more accessible through simple and interactive interfaces. Through this work, MHNfs is now made user-friendly via an interactive interface that allows researchers to perform activity predictions with minimal technical expertise, akin to how LLMs are employed for diverse tasks (see Figure 2).

Prompting-Based Activity Predictions for Low-Data Scenarios. We provide an application that enables the use of the MHNfs model through a simple, interactive user interface based on input prompts. Specifically, we provide a Streamlit application,[44] hosted on HuggingFace[45] at https://huggingface.co/spaces/ml-jku/mhnfs. This interface allows users to perform activity predictions for any target, as long as some active and inactive molecules are known. To do so, the user needs to provide three input prompts, which are a) the molecules of interest for which activity predictions are requested, b) the known active molecules, and c) the known inactive molecules. Conditioned on the input prompts, MHNfs computes activity predictions, which are then presented to the user in an accessible and user-friendly format.

PROBLEM SETTING AND SUMMARY

Few-shot learning in drug discovery addresses the challenge of building bioactivity models using only a small number of molecules with known bioactivities. These bioactivity values may represent activity in a bioassay or against a specific drug target. The available data, referred to as the *support set*, comprise molecular structures paired with binary bioactivity labels. Few-shot models utilize the support set to predict bioactivity for *query molecules*—those with unknown labels. Recent advancements in few-shot activity prediction methods[32,42,43] have demonstrated good performance, effectively identifying meaningful patterns and providing reasonable predictions even in very low-data scenarios, and hold great potential to accelerate drug discovery. However, their adoption is limited due to accessibility barriers. In this work, we address this limitation by providing a simple, interactive user interface that makes these models accessible to practitioners.

METHOD OVERVIEW AND APPLICATION DETAILS

Figure 1 illustrates the virtual screening process in drug discovery following target identification.[46,47] Virtual screening

https://doi.org/10.1021/acs.jcim.4c02373
J. Chem. Inf. Model. 2025, 65, 4243−4250

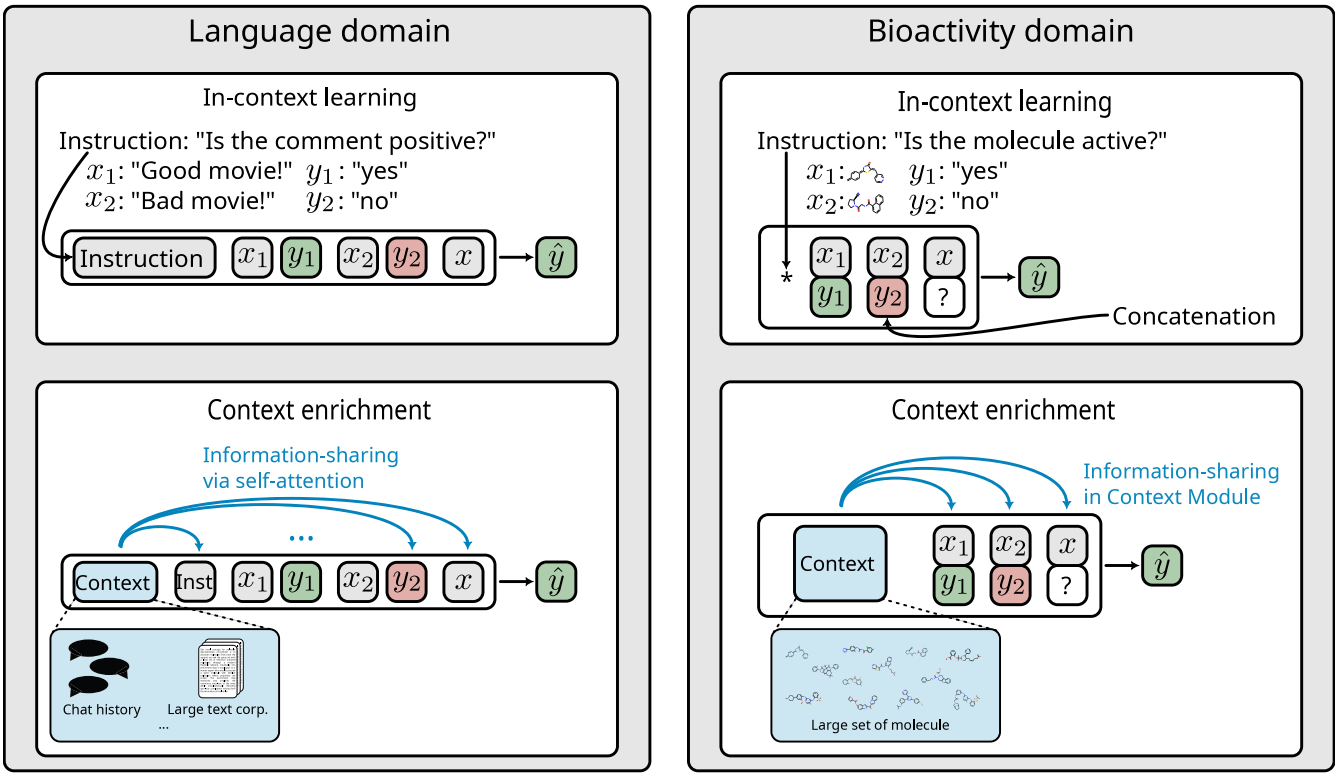


Figure 2. In-context learning and context enrichment. Left: Language domain (adapted from Chen et al.[57] Copyright The Authors. CC BY 4.0). In in-context learning, few-shot tasks are solved by providing an instruction together with known labels and a query x to the LLM. The context window usually includes wider information, e.g., a chat history or a large text corpus, and shares information via self-attention. Right: Bioactivity domain. Analogously, the model is provided with known data (support set) and a query to output activity. Since the instruction does not change for different tasks, it is omitted (indicated by *). `MHNfs` includes wider context, here a large set of training molecules, by retrieving information via the Context Module.

is further broken down into three key substeps: collecting prior knowledge, defining the screening library, and predicting bioactivities. Bioactivity models are a cornerstone of virtual screening, typically trained on data sourced from prior experiments, physical simulations, or text mining. However, in the early stages of drug discovery, the amount of available training data is often severely limited. The proposed application, *MHNfs*, is specifically designed to address this low-data challenge, enabling effective bioactivity predictions during the earliest phases of a drug discovery project.

`MHNfs`: An Embedding-Based Few-Shot Method for Drug Discovery. The backbone of our application is the `MHNfs` model.[42] `MHNfs` consist of three core modules: a) the Context Module, b) the Cross-Attention Module, and c) the Similarity Module.

`MHNfs` takes two inputs: a support set of known active and inactive molecules and a query set of molecules whose activity needs to be predicted. The model first generates an initial low-level representation of the molecules. This representation is refined through the Context and Cross-Attention Modules. Finally, the updated representations are passed to the Similarity Module, which produces the activity predictions for the query set.

Context Module. The Context Module updates input representations by retrieving information from a large set of context molecules, known as the *context set*. This amplifies the covariance structure of the data and removes spurious co-occurrences. The context-enrichment step is conceptually similar to how a large language model (LLM) leverages all available information within its context window (see Section In-Context Learning and Context Enrichment). In `MHNfs`, the context enrichment is realized with a Modern Hopfield Network[48] (Transformer architecture[49]). It performs cross-attention between the molecular input representations and the context set. Notably, the representations of query and support-set molecules are updated independently during this process.

Cross-Attention Module. The Cross-Attention Module enables information sharing between a query and the support-set representations. This information-sharing step, similar to in-context learning, plays a crucial role in embedding-based few-shot methods.[37]

Similarity Module. The Similarity Module computes pairwise similarities between each query molecule and the molecules in the support set. These similarity values are then used as weights in a weighted sum over the support-set labels, ultimately predicting the activity of each query molecule.

Deployed Model with Preconfigured Context. The `MHNfs` backbone was trained and evaluated on the FS-Mol benchmark data set,[42] and its fully trained instance is made accessible through this application. The model is preintegrated with a context set, sampled from the FS-Mol training data, which allows users to perform predictions directly. Consequently, there is no need to retrain the model, adjust its parameters, or manually select a context set; users simply provide the query molecules, along with known active and inactive molecules. For further details on the hyperparameter settings and training procedure, please see Schimunek et al.[42]

https://doi.org/10.1021/acs.jcim.4c02373
J. Chem. Inf. Model. 2025, 65, 4243−4250

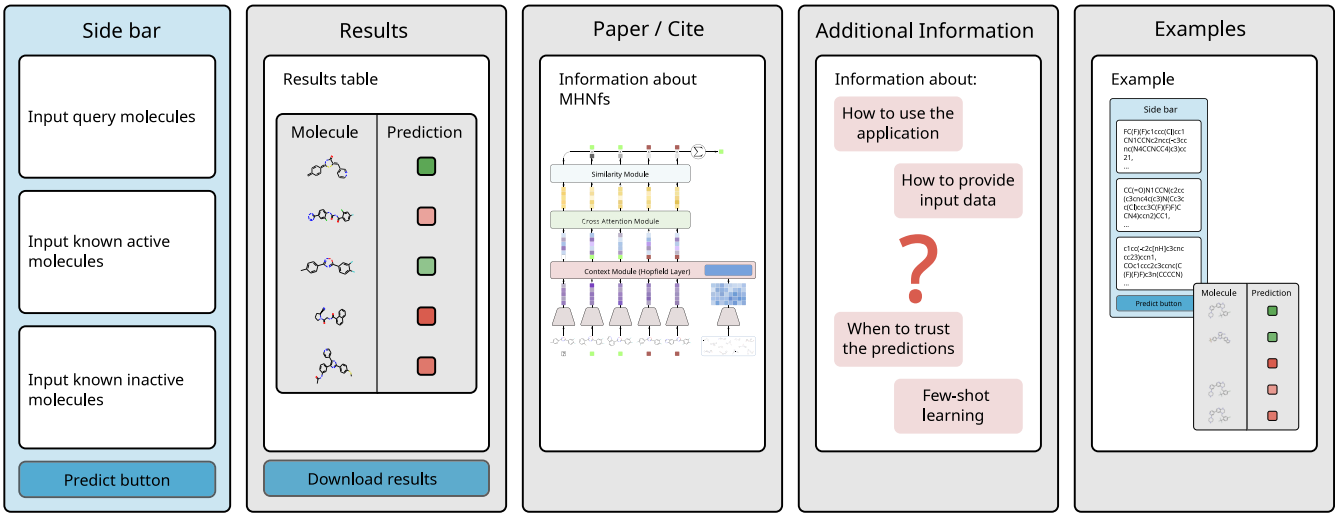


Figure 3. Schematic overview of the Streamlit interface of *MHNfs*. The interface empowers users to input data and execute the pipeline via the sidebar. The main area, highlighted by gray-colored boxes, displays results and additional information. Organized into tabs, represented as separate boxes in the schematic, the main area provides users with an intuitive way to access and explore outputs and supplementary details.

In-Context Learning and Context Enrichment. Following Brown et al.[50] and Dong et al.,[51] **in-context learning** is a paradigm that allows language models to learn tasks given only a few examples in the form of demonstrations. Specifically, a language model is provided with an optional task instruction, known data, and a query sample, and it generates a prediction for the query. In in-context learning, the term context refers to the task-specific information provided to the model as input, including the optional instruction and the known data used to prompt the model.

Long-context LLMs are large language models equipped with an extended context window, enabling them to retrieve and process information from a large corpus,[52,53] such as text chunks from documents or conversation histories in chatbots.[54] This large context window has been shown to significantly enhance performance across various tasks.[55,56] Notably, for downstream tasks, the long context might contain not only task-specific information but also more general information to enrich the encoded tokens with additional information. We refer to this mechanism as **context enrichment**.

For few-shot drug discovery, the in-context learning definition can be easily adopted. In this sense, `MHNfs` is an in-context learning method (see Figure 2). The model is prompted with a support set and a query molecule. The support set includes the known available data for a task in the form of some active and inactive measured molecules. Since for bioactivity models, the task instruction, i.e., "predict activity" is constant for all tasks, it can be omitted. `MHNfs` also includes a context-enrichment mechanism in which broader, more general, not task-specific information is used. The Context Module enriches the vector representations of query and support-set molecules with information retrieved from a large set of additional molecules which were included in the training data.

***MHNfs* HuggingFace Interface.** The HuggingFace[45] application is based on the Streamlit package.[44] It consists of a sidebar and a main area (Figure 3). The sidebar allows users to prompt the model by providing: a) known active molecules, b) known inactive molecules, and c) a set of molecules to predict. These inputs can be entered through text boxes or uploaded as CSV files, with molecules expected in SMILES notation.

The main area includes four tabs: Predictions, Paper/Cite, Additional Information, and Examples. The **Predictions** tab displays results as a table containing queried molecules and their predicted activities, along with a grid plot showing 2D molecular graphs, SMILES strings, and predicted activities. The **Paper/Cite** section links to the publication describing the backbone model of the application. The **Additional Information** section provides details on few-shot learning, `MHNfs`, and guidance for using the application, including a *When to trust the predictions* section that emphasizes staying within the training domain for new tasks. The **Examples** tab offers one example task in which the application performs well and another task in which it is not suitable. This design ensures users can easily interact with the application, access relevant information, and understand its capabilities and limitations.

Model Evaluation. In the original method paper,[42] `MHNfs` was evaluated on the FS-Mol test set. Here, we perform a new experiment and evaluate the *MHNfs* application on different external experimental setups by recasting assay data from PubChem[58] as few-shot scenarios. This evaluates the generalization capabilities of `MHNfs` because the model was originally trained and evaluated on FS-Mol, which is an excerpt from ChEMBL27,[59] and in this work, the model was further evaluated on data from a different source and unseen protein targets without adjusting any of the model parameters.

Review of the FS-Mol Experiments. `MHNfs`[42] was developed and evaluated using the FS-Mol benchmark data set, which comprises 4,938 training tasks and 157 test tasks derived from assay data in ChEMBL27. ChEMBL27[59] is a semimanually curated database of bioactivity values of drug-like molecules—with automated steps to improve data quality[60,61]—that contains approximately 2.4 million compounds and over 15 million activity data points, typically sourced from peer-reviewed scientific publications. On the relatively high-quality FS-Mol test set, which is comparable in quality to the training data, `MHNfs` outperformed other few-shot models, including ADKF-IFT,[41] the IterRefLSTM-based few-shot model,[37] Siamese Networks,[62] Prototypical Net-

https://doi.org/10.1021/acs.jcim.4c02373
J. Chem. Inf. Model. 2025, 65, 4243−4250

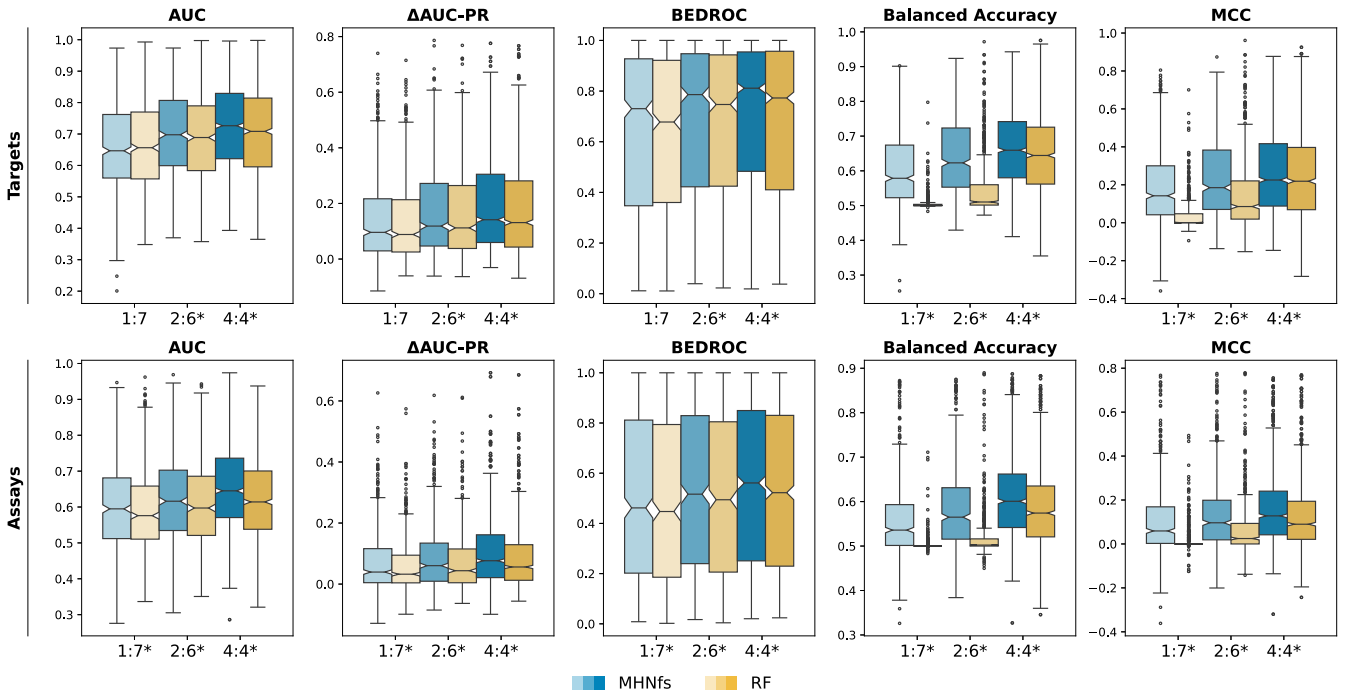


Figure 4. Results on the PubChem-based few-shot experiment with **support-set size 8**. Variation is shown across tasks and three draws of support sets. *n*:*m* indicate a support set with *n* active and *m* inactive molecules. Significantly different performances between the two models are indicated by a star.

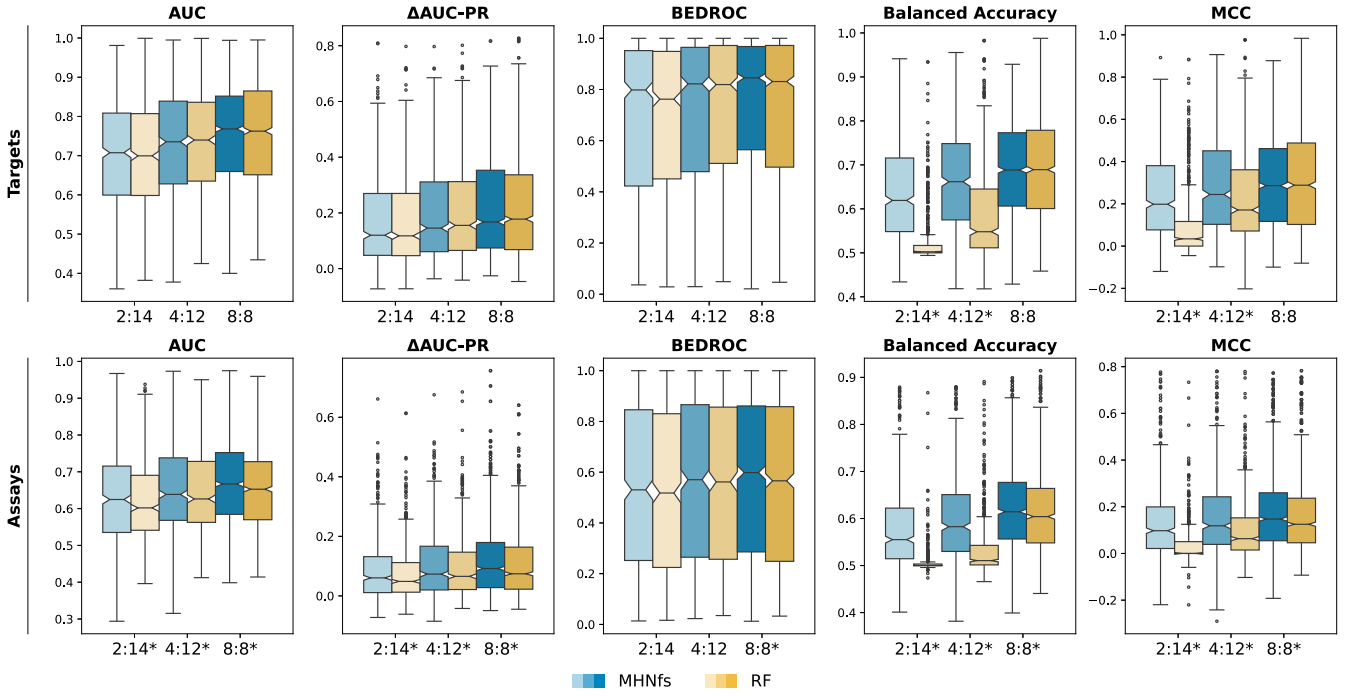


Figure 5. Results on the PubChem-based few-shot experiment with **support-set size 16**. Variation is shown across tasks and three draws of support sets. *n*:*m* indicate a support set with *n* active and *m* inactive molecules. Significantly different performances between the two models are indicated by a star.

works,[63] and Property-Aware Relation Networks.[40] For further details on the FS-Mol test set evaluations and extended results, such as performance across different protein families, please refer to Schimunek et al.[42]

Evaluation on External PubChem-Based Data Sets. Motivation. As we anticipate the application `MHNfs` may also be used in scenarios with reduced data quality, including increased noise and domain shifts compared to the training data, the motivation is to assess the model's robustness. Recent studies[37,42] have indicated that few-shot models can suffer from decreased performance under such conditions. To address this concern, we aim to evaluate `MHNfs`, without adjusting any of the already trained model parameters, using PubChem-based data,[58] a more general database with diverse and potentially

https://doi.org/10.1021/acs.jcim.4c02373
J. Chem. Inf. Model. 2025, 65, 4243−4250

noisier data. Built from over 1,000 data sources, it contains 119 million molecules and 295 million bioactivities, making it significantly larger and more varied, presenting a challenging but important test of MHNfs' adaptability.

Experimental Setup. From PubChem, all single protein assays linked to a protein not present in the FS-Mol training or validation sets were retrieved by comparing UniProt accession keys. The following filtering steps were applied to the retrieved assay data: a) HTS assays with more than 100,000 data points were removed, b) data points with conflicting bioactivity labels were removed, c) molecules were only kept if the MHNfs' preprocessing script could handle the SMILES input, and d) targets with less than 50 active and inactive molecules were removed. Note that for d), we only require the data points for evaluation, while the actual training sets, i.e., support sets, are very small with only 8 or 16 data points. From the filtered data, different experimental setups were built by varying the bioactivity prediction task types, the number of known molecules, and the ratio of known actives and inactives. To consider different bioactivity prediction task types, two different data sets are built that differ from each other by the way the measurements were merged and the tasks were built. For the first data set, the *by-target data set*, bioactivity data linked to the same target were merged, resulting in 229 different tasks. For the second data set, the *by-assay data set*, measurements were merged according to their assay membership and only assays with at least 30 active and 30 inactive molecules were kept, resulting in 127 different tasks. To mimic scenarios with a different number of known molecules and different ratios in terms of actives and inactives, we consider two different support-set sizes, i.e., 8 and 16, and different active and inactive ratios: 1 (2) active and 7 (14) inactive molecules (ratio 1:7), 2 (4) active and 6 (12) inactive molecules (ratio 1:3), and 4 (8) active and 4 (8) inactive molecules (ratio 1:1).

Methods Compared. We compare MHNfs with an RF baseline trained from scratch on the support-set data. The RF baseline was implemented with sklearn using the default hyperparameters. We consider RFs to be a natural default choice for users who want to train models in scenarios with limited data, given their strong performance on small data sets and ease of use.

Metrics. The compared models are evaluated on different metrics: the area under the ROC curve (AUC), the area under the precision-recall curve where the difference to a random classifier is reported (ΔAUC-PR), the Boltzmann-enhanced discrimination of the ROC curve (BEDROC), balanced accuracy, and the Matthews correlation coefficient (MCC).

Results. The results with respect to the different metrics are presented in Figure 4 and Figure 5 for support-set sizes of 8 and 16. The variation of the metrics is shown across tasks and three draws of support sets. The stars on the x-axis labels indicate whether one method significantly outperforms the other (based on a two-sided paired Wilcoxon test). For almost all scenarios and metrics tested, MHNfs significantly outperforms the Random Forest baseline. The median ROC-AUC of MHNfs for targets given only 4 active and 4 inactive molecules, is already above 0.7, which should lead to enrichment of actives in wet-lab testing[64,65] (Figure 4, top left panel).

DISCUSSION AND CONCLUSIONS

In this work, we have introduced an application called *MHNfs*, which provides a simple and accessible user interface to a trained few-shot learning method for bioactivity prediction. With the *MHNfs* application, users can obtain relatively accurate activity predictions for their compound libraries, even if they know as few as one active and one inactive molecule. We validated the application on 229 activity prediction tasks from PubChem, using them to simulate few-shot learning scenarios. The straightforward design, together with the usability and predictive quality of the model, makes few-shot learning broadly accessible to practitioners and researchers beyond experts in chemoinformatics and computer-aided drug discovery. Based on the good performance at very small support sets, e.g., 4 actives and inactives, researchers in the field could be encouraged to start wet-lab screening of molecules early in projects when only scarce information is available. We hope that our work sparks investigations in the design of a context set, whose influence and design has not been thoroughly investigated. Also, for larger support-set sizes, MHNfs has shown promise[42] and can be used without the need to set up model training. However, users should be aware that, as the number of known molecules increases, traditional methods such as RFs may perform better, requiring only moderate additional resources for model training. Snyder et al.[43] report that from 50 measures molecules upward, classic machine learning methods start outperforming few-shot learning methods. We envision that *MHNfs* will become a widely used tool in early stage projects around small molecules and their bioactivities.

Limitations and Recommendations. Few-shot models are known to have limited ability to generalize to unseen domains and to maintain predictive power under domain shifts.[37,42] Therefore, we recommend using this application primarily for protein families—oxidoreductases, kinases, hydrolases, lysases, isomerases, ligases, and translocases—in which MHNfs was specifically evaluated and show good performance. For other targets, users should verify that similar proteins are represented in the FS-Mol training and evaluation sets (see https://github.com/microsoft/FS-Mol/blob/main/datasets/targets/target_info.csv).

ASSOCIATED CONTENT

Data Availability Statement

The HuggingFace application *MHNfs* is available here: https://huggingface.co/spaces/ml-jku/mhnfs. Code and data to train and evaluate the backbone model MHNfs is provided at: https://github.com/ml-jku/MHNfs.

AUTHOR INFORMATION

Corresponding Author

Günter Klambauer − *ELLIS Unit Linz and LIT AI Lab, Institute for Machine Learning, Johannes Kepler University Linz, A-4040 Linz, Austria;* orcid.org/0000-0003-2861-5552; Email: klambauer@ml.jku.at

Authors

Johannes Schimunek − *ELLIS Unit Linz and LIT AI Lab, Institute for Machine Learning, Johannes Kepler University Linz, A-4040 Linz, Austria;* orcid.org/0000-0002-1372-8290

Sohvi Luukkonen − *ELLIS Unit Linz and LIT AI Lab, Institute for Machine Learning, Johannes Kepler University*

https://doi.org/10.1021/acs.jcim.4c02373
J. Chem. Inf. Model. 2025, 65, 4243−4250

Linz, A-4040 Linz, Austria; orcid.org/0000-0001-9387-1427

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.4c02373

Author Contributions

J.S. implemented the methods and algorithms. J.S. and S.L. designed and performed the experiments. J.S., S.L., and G.K. analyzed and interpreted the results. J.S., S.L., and G.K. wrote the manuscript. J.S., S.L., and G.K. conceived and designed the study.

Funding

This research was funded in part by the Austrian Science Fund (FWF) [10.55776/COE12].

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Daniel Kuhn and Lukas Friedrich for their feedback on the *MHNfs* interface. The ELLIS Unit Linz, the LIT AI Lab, the Institute for Machine Learning, are supported by the Federal State Upper Austria. This research was funded in part by the Austrian Science Fund (FWF) [10.55776/COE12]. We thank the projects INCONTROL-RL (FFG-881064), PRIMAL (FFG-873979), S3AI (FFG-872172), DL for GranularFlow (FFG-871302), EPILEPSIA (FFG-892171), FWF AIRI FG 9-N (10.55776/FG9), AI4GreenHeatingGrids (FFG-899943), INTEGRATE (FFG-892418), ELISE (H2020-ICT-2019-3 ID: 951847), Stars4Waters (HORIZON-CL6-2021-CLIMATE-01-01). We thank NXAI GmbH, Audi.JKU Deep Learning Center, TGW LOGISTICS GROUP GMBH, Silicon Austria Laboratories (SAL), FILL Gesellschaft mbH, Anyline GmbH, Google, ZF Friedrichshafen AG, Robert Bosch GmbH, UCB Biopharma SRL, Merck Healthcare KGaA, Verbund AG, GLS (Univ. Waterloo), Software Competence Center Hagenberg GmbH, Borealis AG, TÜV Austria, Frauscher Sensonic, TRUMPF and the NVIDIA Corporation.

■ REFERENCES

(1) Mitchell, M.; Wu, S.; Zaldivar, A.; Barnes, P.; Vasserman, L.; Hutchinson, B.; Spitzer, E.; Raji, I. D.; Gebru, T. Model cards for model reporting. *Proceedings of the conference on fairness, accountability, and transparency*; ACM, 2019; pp 220−229.

(2) Green, D. V. *Machine learning in chemistry: data-driven algorithms, learning systems, and predictions*; ACS Publications, 2019; pp 81−101.

(3) Brown, N.; Ertl, P.; Lewis, R.; Luksch, T.; Reker, D.; Schneider, N. Artificial intelligence in chemistry and drug design. *Journal of Computer-Aided Molecular Design* **2020**, *34*, 709−715.

(4) Tyrchan, C.; Nittinger, E.; Gogishvili, D.; Patronov, A.; Kogej, T. *Computational and data-driven chemistry using artificial intelligence*; Elsevier, 2022; pp 111−159.

(5) Volkamer, A.; Riniker, S.; Nittinger, E.; Lanini, J.; Grisoni, F.; Evertsson, E.; Rodríguez-Pérez, R.; Schneider, N. Machine learning for small molecule drug discovery in academia and industry. *Artificial Intelligence in the Life Sciences* **2023**, *3*, 100056.

(6) Mullard, A. New drugs cost US 2.6 billion to develop. *Nat. Rev. Drug Discovery* **2014**, *13*, 877.

(7) Yoo, J.; Kim, T. Y.; Joung, I.; Song, S. O. Industrializing AI/ML during the end-to-end drug discovery process. *Curr. Opin. Struct. Biol.* **2023**, *79*, 102528.

(8) Bohacek, R. S.; McMartin, C.; Guida, W. C. The art and practice of structure-based drug design: a molecular modeling perspective. *Medicinal research reviews* **1996**, *16*, 3−50.

(9) Ertl, P. Cheminformatics analysis of organic substituents: identification of the most common substituents, calculation of substituent properties, and automatic identification of drug-like bioisosteric groups. *Journal of chemical information and computer sciences* **2003**, *43*, 374−380.

(10) Kirkpatrick, P.; Ellis, C. Chemical space. *Nature* **2004**, *432*, 823−824.

(11) Reymond, J.-L.; Awale, M. Exploring chemical space for drug discovery using the chemical universe database. *ACS chemical neuroscience* **2012**, *3*, 649−657.

(12) Arrowsmith, J. Phase II failures: 2008−2010. *Nat. Rev. Drug Discovery* **2011**, *10* (5), 328−329.

(13) Jayatunga, M. K.; Xie, W.; Ruder, L.; Schulze, U.; Meier, C. AI in small-molecule drug discovery: a coming wave. *Nat. Rev. Drug Discov* **2022**, *21*, 175−176.

(14) Chen, H.; Engkvist, O.; Wang, Y.; Olivecrona, M.; Blaschke, T. The rise of deep learning in drug discovery. *Drug discovery today* **2018**, *23*, 1241−1250.

(15) Walters, W. P.; Barzilay, R. Critical assessment of AI in drug discovery. *Expert opinion on drug discovery* **2021**, *16*, 937−947.

(16) Seidl, P.; Vall, A.; Hochreiter, S.; Klambauer, G. Enhancing activity prediction models in drug discovery with the ability to understand human language. *arXiv preprint arXiv:2303.03363* **2023**.

(17) Mayr, A.; Klambauer, G.; Unterthiner, T.; Steijaert, M.; Wegner, J. K.; Ceulemans, H.; Clevert, D.-A.; Hochreiter, S. Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. *Chemical science* **2018**, *9*, 5441−5451.

(18) Schimunek, J.; Seidl, P.; Elez, K.; Hempel, T.; Le, T.; Noé, F.; Olsson, S.; Raich, L.; Winter, R.; Gokcan, H.; et al. A community effort in SARS-CoV-2 drug discovery. *Molecular Informatics* **2024**, *43*, No. e202300262.

(19) Mayr, A.; Klambauer, G.; Unterthiner, T.; Hochreiter, S. DeepTox: toxicity prediction using deep learning. *Frontiers in Environmental Science* **2016**, *3*, 80.

(20) Fan, W.; Xu, X.; Shen, Y.; Feng, H.; Li, A.; Wang, M. Prediction of protein kinase-specific phosphorylation sites in hierarchical structure using functional information and random forest. *Amino acids* **2014**, *46*, 1069−1078.

(21) Cano, G.; Garcia-Rodriguez, J.; Garcia-Garcia, A.; Perez-Sanchez, H.; Benediktsson, J. A.; Thapa, A.; Barr, A. Automatic selection of molecular descriptors using random forest: Application to drug discovery. *Expert Systems with Applications* **2017**, *72*, 151−159.

(22) Matsumoto, A.; Aoki, S.; Ohwada, H. Comparison of random forest and SVM for raw data in drug discovery: prediction of radiation protection and toxicity case study. *International Journal of Machine Learning and Computing* **2016**, *6*, 145.

(23) Schöning, V.; Krähenbühl, S.; Drewe, J. The hepatotoxic potential of protein kinase inhibitors predicted with Random Forest and Artificial Neural Networks. *Toxicology letters* **2018**, *299*, 145−148.

(24) Babajide Mustapha, I.; Saeed, F. Bioactive molecule prediction using extreme gradient boosting. *Molecules* **2016**, *21*, 983.

(25) Boldini, D.; Grisoni, F.; Kuhn, D.; Friedrich, L.; Sieber, S. A. Practical guidelines for the use of gradient boosting for molecular property prediction. *Journal of Cheminformatics* **2023**, *15*, 73.

(26) Heid, E.; Greenman, K. P.; Chung, Y.; Li, S.-C.; Graff, D. E.; Vermeire, F. H.; Wu, H.; Green, W. H.; McGill, C. J. Chemprop: A machine learning package for chemical property prediction. *J. Chem. Inf. Model.* **2024**, *64*, 9−17.

(27) Fillbrunn, A.; Dietz, C.; Pfeuffer, J.; Rahn, R.; Landrum, G. A.; Berthold, M. R. KNIME for reproducible cross-domain analysis of life science data. *Journal of biotechnology* **2017**, *261*, 149−156.

(28) Ramsundar, B. Molecular machine learning with DeepChem. *Ph.D. thesis*, Stanford University, 2018.

(29) Minnich, A. J.; McLoughlin, K.; Tse, M.; Deng, J.; Weber, A.; Murad, N.; Madej, B. D.; Ramsundar, B.; Rush, T.; Calad-Thomson, S.; et al. AMPL: a data-driven modeling pipeline for drug discovery. *J. Chem. Inf. Model.* **2020**, *60*, 1955−1968.

(30) Turon, G.; Hlozek, J.; Woodland, J. G.; Kumar, A.; Chibale, K.; Duran-Frigola, M. First fully-automated AI/ML virtual screening

cascade implemented at a drug discovery centre in Africa. *Nat. Commun.* **2023**, *14*, 5736.
(31) van den Maagdenberg, H. W.; Šícho, M.; Araripe, D. A.; Luukkonen, S.; Schoenmaker, L.; Jespers, M.; Béquignon, O. J.; González, M. G.; van den Broek, R. L.; Bernatavicius, A. QSPRpred: a Flexible Open-Source Quantitative Structure-Property Relationship Modelling Tool. *Journal of Cheminformatics* **2024**, *16*, 128.
(32) Stanley, M.; Bronskill, J. F.; Maziarz, K.; Misztela, H.; Lanini, J.; Segler, M.; Schneider, N.; Brockschmidt, M. Fs-mol: A few-shot learning dataset of molecules. *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*; NeurIPS, 2021.
(33) Consortium, C. M.; Achdout, H.; Aimon, A.; Alonzi, D. S.; Arbon, R.; Bar-David, E.; Barr, H.; Ben-Shmuel, A.; Bennett, J.; Bilenko, V. A.; et al. Open Science Discovery of Potent Non-Covalent SARS-CoV-2 Main Protease Inhibitors. *BioRxiv* **2020**.
(34) Sun, J.; Jeliazkova, N.; Chupakhin, V.; Golib-Dzib, J.-F.; Engkvist, O.; Carlsson, L.; Wegner, J.; Ceulemans, H.; Georgiev, I.; Jeliazkov, V.; et al. ExCAPE-DB: an integrated large scale dataset facilitating Big Data analysis in chemogenomics. *Journal of Cheminformatics* **2017**, *9*, 41.
(35) Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; Pande, V. MoleculeNet: a benchmark for molecular machine learning. *Chemical science* **2018**, *9*, 513−530.
(36) Sturm, N.; Mayr, A.; Le Van, T.; Chupakhin, V.; Ceulemans, H.; Wegner, J.; Golib-Dzib, J.-F.; Jeliazkova, N.; Vandriessche, Y.; Böhm, S.; et al. Industry-scale application and evaluation of deep learning for drug target prediction. *Journal of Cheminformatics* **2020**, *12*, 26.
(37) Altae-Tran, H.; Ramsundar, B.; Pappu, A. S.; Pande, V. Low data drug discovery with one-shot learning. *ACS central science* **2017**, *3*, 283−293.
(38) Adler, T.; Brandstetter, J.; Widrich, M.; Mayr, A.; Kreil, D.; Kopp, M.; Klambauer, G.; Hochreiter, S. Cross-domain few-shot learning by representation fusion. *arXiv preprint arXiv:2010.06498* **2020**.
(39) Guo, Z.; Zhang, C.; Yu, W.; Herr, J.; Wiest, O.; Jiang, M.; Chawla, N. V. Few-shot graph learning for molecular property prediction. *Proceedings of the web conference* **2021**, 2559−2567.
(40) Wang, Y.; Abuduweili, A.; Yao, Q.; Dou, D. Property-aware relation networks for few-shot molecular property prediction. *Advances in Neural Information Processing Systems* **2021**, *34*, 17441−17454.
(41) Chen, W.; Tripp, A.; Hernández-Lobato, J. M. Meta-learning Adaptive Deep Kernel Gaussian Processes for Molecular Property Prediction. *Eleventh International Conference on Learning Representations*; ICLR, 2023.
(42) Schimunek, J.; Seidl, P.; Friedrich, L.; Kuhn, D.; Rippmann, F.; Hochreiter, S.; Klambauer, G. Context-enriched molecule representations improve few-shot drug discovery. *Eleventh International Conference on Learning Representations*; 2023.
(43) Snyder, S. H.; Vignaux, P. A.; Ozalp, M. K.; Gerlach, J.; Puhl, A. C.; Lane, T. R.; Corbett, J.; Urbina, F.; Ekins, S. The Goldilocks paradigm: comparing classical machine learning, large language models, and few-shot learning for drug discovery applications. *Communications Chemistry* **2024**, *7*, 134.
(44) Streamlit • A faster way to build and share data apps — streamlit.io. https://streamlit.io/ [Accessed 24-04-2024].
(45) HuggingFace − The AI community building the future. — huggingface.co. https://huggingface.co/ [Accessed 24-04-2024].
(46) Ain, Q. u.; Batool, M.; Choi, S. TLR4-targeting therapeutics: structural basis and computer-aided drug discovery approaches. *Molecules* **2020**, *25*, 627.
(47) Sun, D.; Gao, W.; Hu, H.; Zhou, S. Why 90% of clinical drug development fails and how to improve it? *Acta Pharmaceutica Sinica B* **2022**, *12*, 3049−3062.
(48) Ramsauer, H.; Schäfl, B.; Lehner, J.; Seidl, P.; Widrich, M.; Adler, T.; Gruber, L.; Holzleitner, M.; Pavlović, M.; Sandve, G. K.; et al. Hopfield networks is all you need. *arXiv preprint arXiv:2008.02217* **2020**.
(49) Vaswani, A. Attention is all you need. *Advances in Neural Information Processing Systems*; NIPS, 2017.
(50) Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. *Advances in neural information processing systems* **2020**, *33*, 1877−1901.
(51) Dong, Q.; Li, L.; Dai, D.; Zheng, C.; Wu, Z.; Chang, B.; Sun, X.; Xu, J.; Sui, Z. A survey on in-context learning. *arXiv preprint arXiv:2301.00234* **2022**.
(52) Borgeaud, S.; Mensch, A.; Hoffmann, J.; Cai, T.; Rutherford, E.; Millican, K.; Van Den Driessche, G. B.; Lespiau, J.-B.; Damoc, B.; Clark, A.; et al. Improving language models by retrieving from trillions of tokens. *International conference on machine learning*; ICML, 2022; pp 2206−2240.
(53) An, S.; Ma, Z.; Lin, Z.; Zheng, N.; Lou, J.-G. Make Your LLM Fully Utilize the Context. *arXiv preprint arXiv:2404.16811* **2024**.
(54) Chen, L.; Liu, Z.; He, W.; Li, Y.; Luo, R.; Yang, M. Long Context is Not Long at All: A Prospector of Long-Dependency Data for Large Language Models. *arXiv preprint arXiv:2405.17915* **2024**.
(55) Bai, Y.; Lv, X.; Zhang, J.; Lyu, H.; Tang, J.; Huang, Z.; Du, Z.; Liu, X.; Zeng, A.; Hou, L.; et al. Longbench: A bilingual, multitask benchmark for long context understanding. *arXiv preprint arXiv:2308.14508* **2023**.
(56) Zhang, L.; Li, Y.; Liu, Z.; Liu, J.; Yang, M. Marathon: A race through the realm of long context with large language models. *arXiv preprint arXiv:2312.09542* **2023**.
(57) Chen, Y.; Zhong, R.; Zha, S.; Karypis, G.; He, H. Meta-learning via Language Model In-context Tuning. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*; Association for Computational Linguistics, 2022; pp 719−730, DOI: .
(58) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; et al. PubChem 2023 update. *Nucleic acids research* **2023**, *51*, D1373−D1380.
(59) Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M. P.; Mosquera, J. F.; Mutowo, P.; Nowotka, M.; et al. ChEMBL: towards direct deposition of bioassay data. *Nucleic acids research* **2019**, *47*, D930−D940.
(60) Papadatos, G.; Gaulton, A.; Hersey, A.; Overington, J. P. Activity, assay and target data curation and quality in the ChEMBL database. *Journal of computer-aided molecular design* **2015**, *29*, 885−896.
(61) Zdrazil, B.; Felix, E.; Hunter, F.; Manners, E. J.; Blackshaw, J.; Corbett, S.; de Veij, M.; Ioannidis, H.; Lopez, D. M.; Mosquera, J. F.; et al. The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic acids research* **2024**, *52*, D1180−D1192.
(62) Koch, G.; Zemel, R.; Salakhutdinov, R. Siamese neural networks for one-shot image recognition. *Proceedings of the 32 nd International Conference on Machine*; ICML, 2015; pp 1−30.
(63) Snell, J.; Swersky, K.; Zemel, R. Prototypical networks for few-shot learning. *Advances in neural information processing systems*; NIPS, 2017; p *30*.
(64) Simm, J.; Klambauer, G.; Arany, A.; Steijaert, M.; Wegner, J. K.; Gustin, E.; Chupakhin, V.; Chong, Y. T.; Vialard, J.; Buijnsters, P.; et al. Repurposing high-throughput image assays enables biological activity prediction for drug discovery. *Cell chemical biology* **2018**, *25*, 611−618.
(65) Hofmarcher, M.; Rumetshofer, E.; Clevert, D.-A.; Hochreiter, S.; Klambauer, G. Accurate prediction of biological assays with high-throughput microscopy images and convolutional networks. *J. Chem. Inf. Model.* **2019**, *59*, 1163−1171.

https://doi.org/10.1021/acs.jcim.4c02373
J. Chem. Inf. Model. 2025, 65, 4243−4250