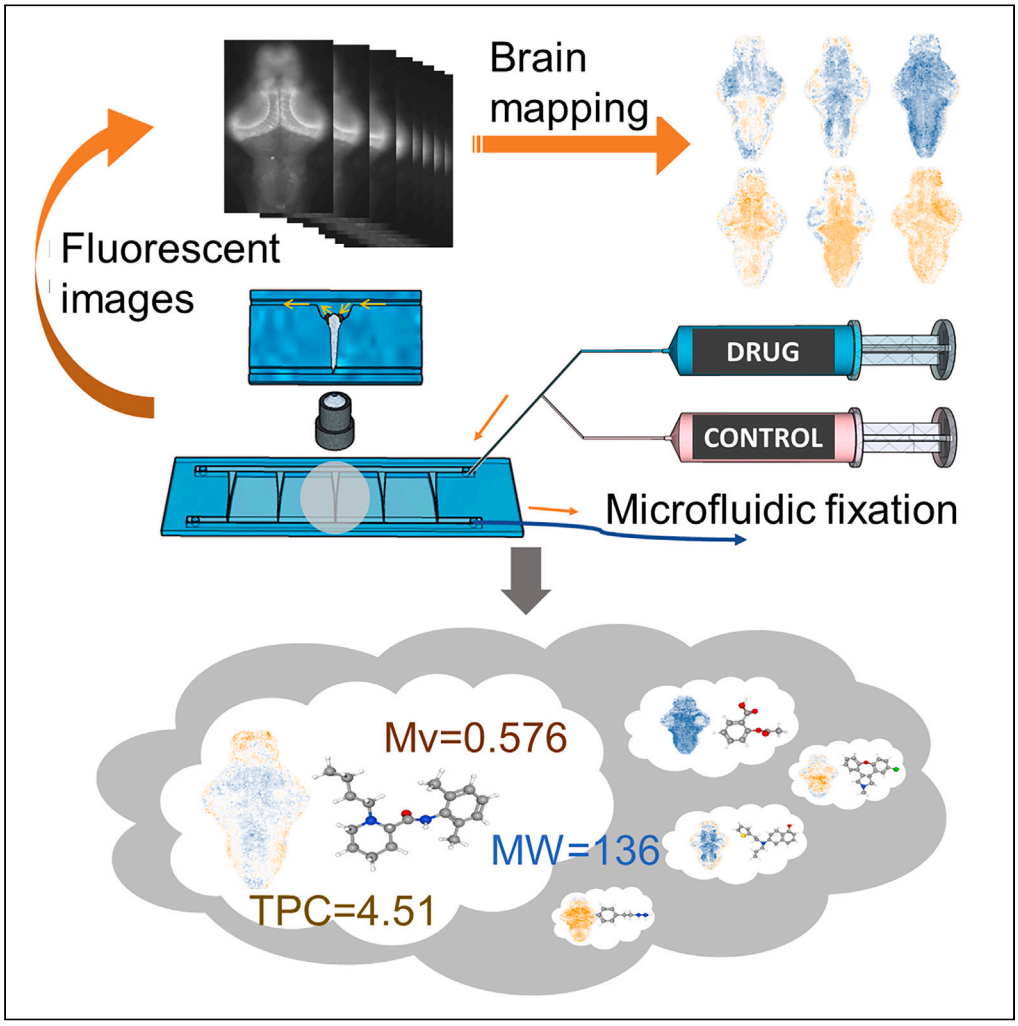
iScience

CellPress
OPEN ACCESS

Article

Few-shot meta-learning applied to whole brain activity maps improves systems neuropharmacology and drug discovery



Xuan Luo, Yanyun Ding, Yi Cao, ..., Xin Wang, Jin Zhang, Peng Shi

zhangj9@sustech.edu.cn (J.Z.)
pengshi@cityu.edu.hk (P.S.)

Highlights

Integrating few-shot meta-learning with brain activity mapping for drug discovery

Utilization of high-throughput whole-brain activity mapping for pharmaceutical repurposing

Limited brain physiology data for fast and effective learning of pharmacological features

Successful identification of drug candidates by an advanced "learning to learn" approach

Luo et al., iScience 27, 110875
October 18, 2024
https://doi.org/10.1016/j.isci.2024.110875

iScience

CellPress
OPEN ACCESS

Article

Few-shot meta-learning applied to whole brain activity maps improves systems neuropharmacology and drug discovery

Xuan Luo,[1,2,3,10] Yanyun Ding,[2,9,10] Yi Cao,[4] Zhen Liu,[1] Wenchong Zhang,[1] Shangzhi Zeng,[2] Shuk Han Cheng,[4] Honglin Li,[5] Stephen J. Haggarty,[6] Xin Wang,[7] Jin Zhang,[2,3,*] and Peng Shi[1,2,8,11,*]

SUMMARY

In this study, we present an approach to neuropharmacological research by integrating few-shot meta-learning algorithms with brain activity mapping (BAMing) to enhance the discovery of central nervous system (CNS) therapeutics. By utilizing patterns from previously validated CNS drugs, our approach facilitates the rapid identification and prediction of potential drug candidates from limited datasets, thereby accelerating the drug discovery process. The application of few-shot meta-learning algorithms allows us to adeptly navigate the challenges of limited sample sizes prevalent in neuropharmacology. The study reveals that our meta-learning-based convolutional neural network (*Meta*-CNN) models demonstrate enhanced stability and improved prediction accuracy over traditional machine-learning methods. Moreover, our BAM library proves instrumental in classifying CNS drugs and aiding in pharmaceutical repurposing and repositioning. Overall, this research not only demonstrates the effectiveness in overcoming data limitations but also highlights the significant potential of combining BAM with advanced meta-learning techniques in CNS drug discovery.

INTRODUCTION

Neurological disorders represent one of the most formidable challenges to global health, with their impact extending to more than a billion individuals across the world. According to the report released by the World Health Organization (WHO), the global incidence of depression has reached an alarming level, with over 264 million individuals recorded till 2018.[1] While more severe disorders, such as Alzheimer's disease (AD) and schizophrenia, engendering larger psychological and physical detriment on both patients and their families. As the primary therapeutic option for neurological diseases, drug treatment has provided hope to patients on curing or alleviating. However, compared with other therapeutic areas, the success rate in central nervous system (CNS) drug development showed a considerably lower trend, In the sphere of clinical development, a mere 7% of drugs designed for the CNS ultimately find their way to the consumer market,[2] substantially lower than the average 15% across all other therapeutic areas. A significant contributing factor to this low success rate is the inherent complexity of brain physiology, which features a myriad of specialized cells operating within a highly intricate architectural framework. Over numerous years, a multitude of studies have thoroughly explored the biological mechanisms underpinning various CNS disorders, placing particular emphasis on intermolecular interactions, membrane receptors, ion channels, and regulatory pathways.[2,3] Yet, our comprehension of brain diseases remains largely elusive, and the translation of CNS research findings into clinically viable treatments has proved immensely challenging. For drug screening, there is a growing consensus that many traditional *in vitro* systems are inadequate in bridging the gap between overly simplified models and the complexity of the brain organ. Additionally, the presence of the blood-brain barrier (BBB), which regulates the penetration of most therapeutic agents, poses a significant pharmacokinetic hurdle to CNS drugs.[4,5] Translating CNS disease findings into clinically effective drugs is still challenging and requires overcoming these various hurdles.[2]

[1]Department of Biomedical Engineering, City University of Hong Kong, Kowloon 999077, Hong Kong SAR, China
[2]National Center for Applied Mathematics Shenzhen, Shenzhen 518000, China
[3]Department of Mathematics, Southern University of Science and Technology, Shenzhen 518055, China
[4]Department of Biomedical Sciences, City University of Hong Kong, Kowloon 999077, Hong Kong SAR, China
[5]Innovation Center for AI and Drug Discovery, East China Normal University, Shanghai 200062, China
[6]Chemical Neurobiology Laboratory, Precision Therapeutics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Department of Neurology, Harvard Medical School, Boston, MA 02114, USA
[7]Department of Surgery, Chinese University of Hong Kong, Kowloon 999077, Hong Kong SAR, China
[8]Shenzhen Research Institute, City University of Hong Kong, Shenzhen 518057, China
[9]Institute of Applied Mathematics, Shenzhen Polytechnic University, Shenzhen 518055, China
[10]These authors contributed equally
[11]Lead contact
*Correspondence: zhangj9@sustech.edu.cn (J.Z.), pengshi@cityu.edu.hk (P.S.)
https://doi.org/10.1016/j.isci.2024.110875

In the past, great efforts have been made to enhance the efficiency of CNS-targeted drug discovery based on the understanding of neuropharmacology.[6–8] However, more than 90% of the candidate compounds fail to progress beyond preclinical stages.[9] With the development of high-throughput techniques and the growth of data volumes, machine learning methodologies are increasingly employed across diverse areas of biological research. Examples of such applications include genomic element identification and classification,[10,11] predictions of protein binding,[12,13] and identification of key transcriptional drivers of cancer.[14,15] Deep learning approaches are particularly advantageous in their ability to identify predictive patterns in complex biological systems. However, these methods require vast amounts of data and computational resources, making it challenging to train models with limited experimental data,[16,17] while generating extensive and well-labeled datasets is time-consuming and unfeasible for certain biological tasks. Meta-learning is an emerging technique that allows models to learn from prior experiences and improve their learning performance on related tasks.[18,19] This algorithm exhibits a remarkable aptitude for few-shot learning, setting itself apart from transfer learning and automated machine learning (AutoML) by inherently learning new tasks using prior knowledge. Few-shot meta-learning excels in rapidly adapting to new tasks with minimal training examples, which is highly advantageous in domains where data are limited or costly to obtain. This capability is particularly pertinent in drug discovery, where each new drug represents a unique scenario. By learning to generalize from a small number of examples with high quality of the meta-training set, meta-learning algorithms can effectively address new, unseen tasks. In comparison, transfer learning leverages knowledge from related tasks to enhance performance on the target task. It is most effective when the source and target tasks share similarities, substantial differences between tasks can lead to suboptimal performance due to the inapplicability of transferred features.[20] Additionally, fine-tuning pre-trained models necessitates meticulous hyperparameter tuning to prevent overfitting, especially when the target dataset is small. Reinforcement learning (RL) focuses on learning optimal policies through interactions with an environment to maximize cumulative rewards.[21] It typically requires numerous interactions to develop effective policies, making it computationally intensive and time-consuming.[22] Therefore, the meta-learning approach, is more suited for static classification tasks such as drug discovery, where rapid adaptation to new data are essential. In the applications of biomedical areas, meta-learning is increasingly being used to improve computational efficiency and reduce the required amount of experimental data.[23]

In our previous study,[24] we demonstrated a microfluidics-based system can execute a high-throughput, *in vivo* drug screen that combines automated whole-brain activity mapping with unsupervised learning. This combination enabled the successful prediction of therapeutic uses for anti-epileptic drugs (Anatomical Therapeutic Chemical (ATC) code: N03). Compared with other commonly used drug discovery strategies, such as recombinant DNA technology and tissue culture techniques, the early stages of pharmaceutical discovery often involve the use of recombinant proteins or heterologous cellular models.[25] These methods allow for the detailed study of specific biological targets and the screening of compounds in a controlled environment. More recently, screening methods using organism models have been developed.[26] These platforms offer the advantage of evaluating drug effects in the context of whole-organism biology, providing insights into organ-specific physiology and potential systemic effects. However, these methods are limited in their ability to perform large-scale chemical screens. The complexity and variability inherent in whole-organism models present significant challenges for high-throughput screening, which is essential for the rapid identification of potential therapeutic compounds. In this high-throughput *in vivo* drug screening strategy, BAM data serve as evaluation inputs, providing the potential to directly correlate the brain physiology of zebrafish with the clinical potency of compounds. This innovative use of BAM data allow us to bypass traditional reliance on chemical structure or single molecular targets, instead generating drug assessments purely based on physiological phenotypes, thus providing a more accurate and physiologically relevant assessment of drug effects. Our method, which involves screening CNS-active compounds using BAM, creates a foundational dataset for systems neuropharmacology.

In this study, we have expanded the dataset and explored its application with advanced machine learning and other computational approaches to improve the process of CNS drug discovery and repurposing. Existing data-driven drug discovery methods often rely on algorithms like support vector machines,[27] random forests,[28] and k-nearest neighbors.[29] However, these approaches are typically model-free or single-layer models, which can lead to overfitting when dealing with small datasets. Moreover, the current number of approved CNS drugs does not suffice for traditional machine learning requirements. To address these challenges, we adopted a few-shot learning framework based on the meta-learning paradigm.[30–32] Studies have demonstrated the effectiveness of few-shot methods in decoding cognitive states from brain activity maps[33] and have shown that meta-learning significantly improves the accuracy of decoding electroencephalogram (EEG) signals compared to traditional and transfer learning methods, particularly in noisy data environments.[34] Inspired by these methods, our integration of high-throughput BAM technology with few-shot meta-learning represents a significant advance in drug discovery. The results show that convolutional neural network (*Meta*-CNN) model gives nearly 58% improvement in prediction accuracy (from 1/7 to 72.5%) with BAM input for the identification of potent anti-Parkinson leads, suggesting that few-shot learning-based processing of the BAM database is advantageous for developing CNS therapeutic agents, with substantially reduced amount of experimental data, and the combination of high-throughput BAM technology, and few-shot learning presents a promising approach for drug discovery. By providing a comprehensive brain physiology library of zebrafish and addressing the limitations of existing algorithms, our study offers a robust framework for discovering and repurposing CNS therapeutic agents. This establishes a potential paradigm for drug assessment and development.

RESULTS

System neuropharmacology by machine learning-based BAMing

Developing pathophysiology-related models for pre-clinical CNS drug discovery remains challenging.[2,35] We aim to expedite the process by combining high-throughput whole-animal physiology phenotyping with meta-learning. The experimental setup relies on a largescale *in vivo*

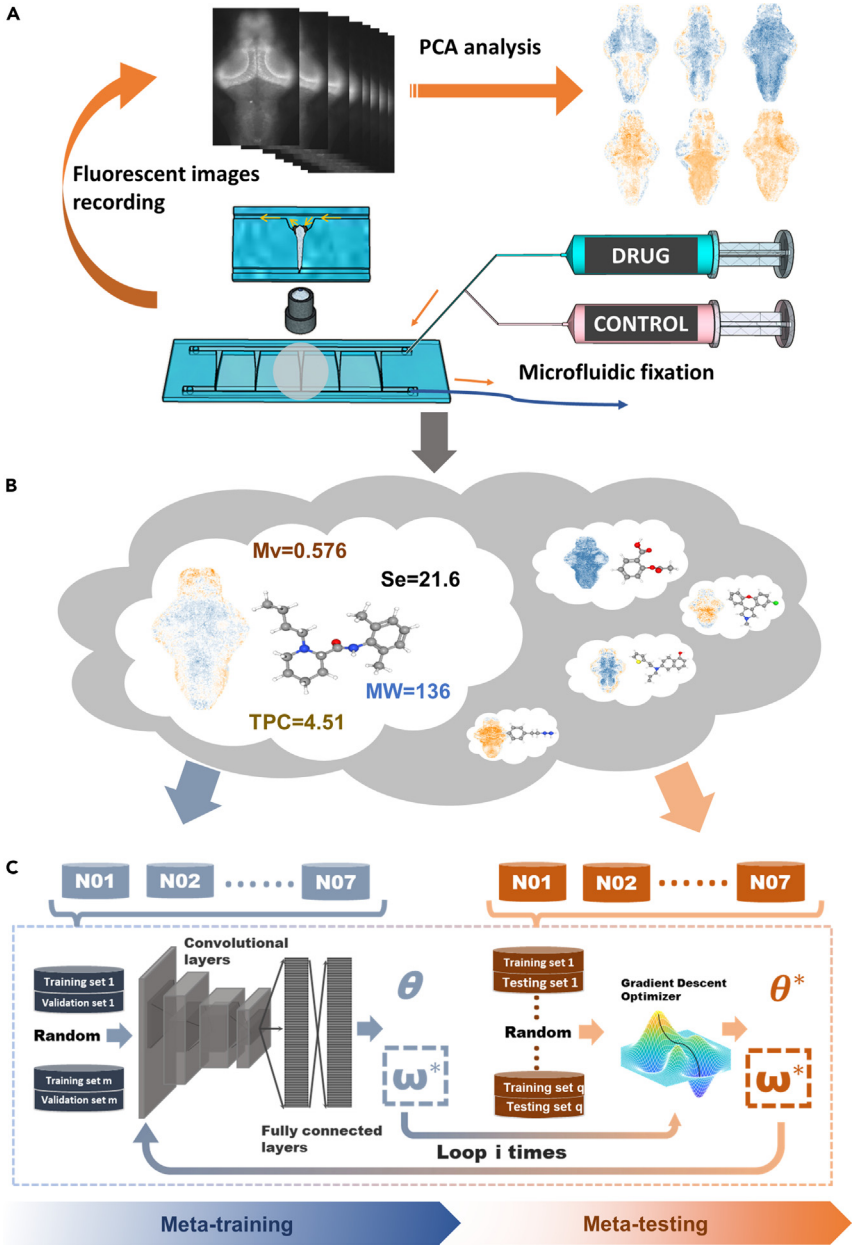


Figure 1. Structure of neuropharmacology analysis methods

(A) Illustration of high-throughput brain activity mapping (BAM) technology in zebrafish larvae.

(B) A structured database designed to inform and train machine learning algorithms.

(C) Schematic representation of a two-phase meta-learning framework. In the meta-training phase, the dataset is partitioned into a training set and a validation set (repeated m times), a multi-layer neural network model was trained to ascertain the optimal hyperparameters ω*. In the meta-testing phase, the dataset undergoes a random split into training and testing subsets (repeated q times), it stabilizes the hyperparameters ω* and trains the model parameters Θ*.

drug screening strategy, which allowed for whole-brain activity imaging and subsequent analysis of drug-induced neuronal response to provide a comprehensive understanding of complex therapeutic effects (Figure 1A).[6,24] To resolve the spatial neuronal activity associated with different drug treatments across a whole brain, we used transgenic zebrafish (elavl3:GCaMP5G) engineered with calcium indicators and a microfluidic system previously developed by our team.[36] In the context of CNS drug discovery, the scarcity of data are evident, neuropharmacology often faces the challenge of limited data due to the high cost and complexity of generating brain activity data. WHO has published only 637 CNS drugs in their dictionary, with the quantity for ATC: N04: anti-Parkinson drugs being particularly small at just thirty-eight. Furthermore, only 220 of these CNS drugs have been approved by the Food and Drug Administration (FDA) (Table S1). This limited number of

Table 1. 3 network configurations for average final testing accuracy (*n* = 6): Random Forest (abbreviated as RF in figures) model, CNN-only (abbreviated as CNN in figures) model, and *Meta*-CNN learning algorithm (shot means the sample size)

Input	Label	Algorithm	Training	Validation	Training	Testing	Average accuracy
BAMs	ATC	Random forest	60%	–	–	40%	26.5%
BAMs	ATC	CNN	18-shot	–	–	3-shot	27.8%
BAMs	ATC	*Meta*-CNN	4-shot	3-shot	11-shot	3-shot	56.4%

available drugs makes it infeasible to train traditional machine learning models effectively. Our approach leverages few-shot meta-learning, which is specifically designed to be used with limited data (Figures 1B and 1C). By using meta-learning, our method can generalize a learning model from a small number of samples, making it highly suitable for neuropharmacological applications given the data scarcity. In this study, beyond employing few-shot learning, we also developed a two-layer optimization model to effectively reduce the risk of overfitting (detailed in STAR Methods), which is a common issue with traditional machine learning models when handling small datasets, thus improving the reliability and robustness of predictions in small sample scenario.[37] The potential impacts of our research extend beyond accelerating CNS drug discovery, as the framework could also be applicable to other fields facing similar data limitations, offering a robust solution for performing reliable and accurate predictions with limited data. By applying the "learning to learn" strategy to analyze the BAM dataset, we can train a model across a plethora of learning tasks and resolve new learning tasks using small training samples. This synergistic integration of machine learning algorithms with high-throughput brain activity mapping shows the promise to break the limitations in current drug discovery and unlock frontiers in neuropharmacology studies.

Pharmacological prediction using BAM analysis

Extended from our previously published research,[24] the processing of the BAM dataset was divided into two steps: image pre-processing and BAM clustering via unsupervised learning (Figures S1 and S2). From the clustering results, we found that BAM-cluster 4 exhibited significant overlap with antiepileptics anatomical therapeutic chemical (ATC categories N03), which generally shows spatially distributed inhibition of neuronal activities across different parts of a whole larval brain (Figure S3). As practiced, we found that it was challenging to extract additional drug-indicative features from other clusters. One of the possible reasons could be the heavy dependence of the scale of the dataset for unsupervised learning. To address this issue, we enriched the size of the database (Tables S1 and S2) and tried to integrate certain supervised learning components to analyze the BAM features to pharmacological dissection. We used the popular CNN framework as a base model, to optimize the learning outcome, we experimented with various data partition strategies and determined that, after 200 iterations, a smaller testing dataset was more conducive to learning (Table S3). Nevertheless, the percentage improvement in delivering accurate output remained minimal due to the constraints posed by the relatively small scale of the BAM library size. We then incorporated meta-learning, algorithm demonstrated to be effective in managing few-shot learning problems, into the CNN model. As a result, by refining the data partition (Table S3), we observed a significant increase in predictive accuracy for output labels, showing approximately 20% improvement compared with CNN-only model, the results of multi-experiments even proved a stable 30% increase (Table 1). This indicates not only an improved classification success rate during the testing phases of each iteration but also an overall increase in the learning efficiency of the network. In order to provide a comprehensive evaluation of model performance, we applied several evaluation metrics. The F1 score is particularly useful in situations with an uneven class distribution, which is applicable to our dataset.[38] The area under the curve (AUC) measures the entire two-dimensional area underneath the receiver operating characteristic (ROC) curve, providing a single scalar value to compare models.[39] Additionally, we discussed Cohen's Kappa, a statistical measure that compares observed accuracy with expected accuracy, accounting for the possibility of agreement occurring by chance.[40] From random data separation in the *Meta*-CNN experiments (n = 6), we found that after 200 iterations, the average accuracy improved from 19.7% to 56.4% (Figures 2A and 2B). This indicates that the classification ability of the *Meta*-CNN model improved significantly through learning. The F1 score also improved from 12.4% to 56.5%, demonstrating the model's ability to attain both good precision and recall. The AUC value, indicating low predictive power, increased from 53.2% to 74.6%, suggesting a good discrimination ability. Cohen's Kappa improved from 6.3% to 49.1%, surpassing the threshold for moderate agreement, which indicates that our learning models' accuracy improved for classification tasks. More detailed information on the confusion matrix and metrics' curves among different categories also supports this conclusion (Figures S4 and S5). Almost all categories of CNS drugs showed improved classification performance after *Meta*-CNN model learning, especially N04: anti-Parkinson, which achieved an average true positive (TP) rate of 72.4% across six experiments and showed significant improvements in all the evaluation metrics.

We also included the classical machine learning method, random forest, in our comparison. Random forest operates by creating numerous decision trees, collectively forming what is known as random forests, to tackle classification or regression tasks.[41] Random forest demonstrates commendable performance when handling datasets of moderate size and high dimensionality, showing resilience in addressing missing values and outliers. However, it may still succumb to overfitting in scenarios such as the presence of noise or outliers in the dataset, or when there is a high degree of correlation between features in the training data.[42] We firstly determined the optimal hyperparameters for the random forest model using a grid search test. The chosen values, listed in Table S4, were selected based on commonly used values for classification tasks.[43,44] Besides the hyperparameters, the size of the dataset for testing also affected the classification results. A test

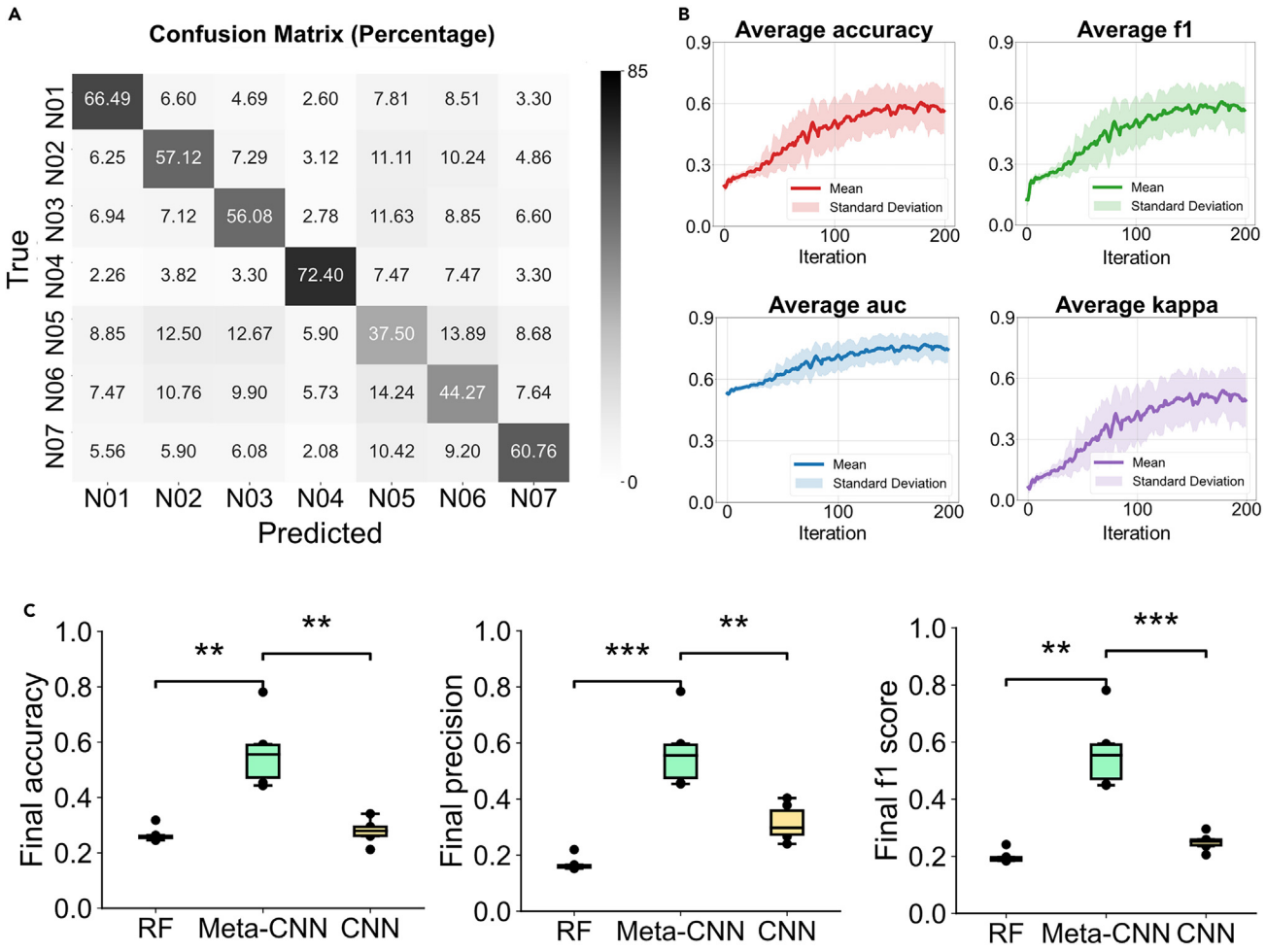


Figure 2. Results for learning models comparison

(A) The average confusion matrix from Meta-CNN experiments ($n = 6$).

(B) The average performance curves of key metrics: accuracy, F1 score, AUC, and Cohen's Kappa, showing the improvement of these parameters over 200 iterations.

(C) The boxplots comparing Random Forest, Meta-CNN, and CNN models. From left to right, the metrics shown are final accuracy, final precision, and final F1 score values. Data are represented as boxplots, including median and interquartile range (IQR), with whiskers extending to 1.5 times the IQR, ** indicates $p < 0.01$, and *** indicates $p < 0.001$ by t test.

size of 40% was found to be optimal according to our results (Figure S6). We conducted six experiments on the bam-ATC data using the optimized random forest model. The results consistently showed that the prediction accuracy for ATC codes: N01-N04 and N07 was low, while the average TP rate for N05 and N06 was around 51.1% and 46.9%, respectively (Figures S6A and S6B). This discrepancy is attributed to the excessively large sample sizes for N05 and N06 compared to the significantly smaller sample sizes for other categories. It indicates that the random forest algorithm is inadequate for this small sample dataset with highly correlated data features. Metrics such as accuracy, F1 score, and precision also demonstrated the insufficiency of the random forest model for small datasets (Figure S6C). In the analysis of the CNN model ($n = 6$), it showed a weak but observable classification ability, with a 43.3% TP rate for N04: anti-Parkinson (Figure S6D). Evaluation metrics revealed a trend where values increased during the early iterations but decreased later, indicating a weak classification ability of the CNN model on the ATC-BAM dataset (Figure S6E). Further analysis comparing random forest, *Meta*-CNN, and CNN models showed that the *Meta*-CNN demonstrated significantly better performance not only in accuracy but also in precision and F1 score values (Table 2; Figure 2C). For more detailed statistical analyses to support our claims of improved performance and stability, we added the additional analysis of statistical metrics, including appropriate hypothesis testing and effect size calculations. The t test is a statistical measure used to determine whether there is a significant difference between the means of two groups.[45] The Wilcoxon signed-rank test is a non-parametric statistical test used to compare two related samples. It assesses whether their population mean ranks differ, making it a useful alternative to the paired t test when the data do not meet the assumptions of normality.[46] Cohen's d is a measure of effect size that quantifies the difference between two group means in terms of standard deviation units. It is a standardized measure that allows for the comparison of effect sizes across

Table 2. The final accuracy comparison of Random Forest and *Meta*-CNN models, and the comparison of CNN and *Meta*-CNN models using t test, Wilcoxon test, and Cohen's d effect size

Comparison	t test *p* value	Wilcoxon *p* value	Cohen's d Effect Size
RF vs. *Meta*-CNN	3.26E-03	3.13E-02	−3.34
CNN vs. *Meta*-CNN	1.36E-03	3.13E-02	3.09

different studies and contexts.[47] We conducted statistical analyses comparing the performance metrics of the random forest, *Meta*-CNN, and CNN model, as shown in Figure 2C, the results indicate that the *Meta*-CNN model exhibits significantly better performance not only in terms of accuracy but also in precision and F1 score values after training.

To ensure that these results were not circumstantial, we randomly divided our BAM library into different data partitions for 9 trials and plotted the Uniform Manifold Approximation and Projection (UMAP)[48] for the final 50 iterations results for both *Meta*-CNN and CNN models (Figures 3A and 3B). In the CNN-only configuration, with the same convolutional layers and fully connected layer configuration as *Meta*-CNN, after 200 iterations, the mean predictive accuracy (n = 9) improved from 14.3% to 35.9%, marking an increase of 21.6%. In the *Meta*-CNN model, the average testing accuracy (n = 9) improved from 14.3% to 56.7% with a 42.4% increase (Figure 3C). Through a closer look at the variation in accuracy trends over the iterations, we noted that the testing accuracy in the CNN-only model peaked around the 30th iteration before descending, while in the *Meta*-CNN model, the predictive accuracy displayed a steady upward trajectory, highlighting a slow yet persistent learning after the integration of the meta-learning algorithm (Figures 3C and 3D). For the maximum predictive accuracy across the learning curves (n = 9), for mean values of seven CNS categories, the *Meta*-CNN network surpassed the CNN-only network in achieving the maximum predictive accuracy (Figure 3E, detailed in Table S5). To evaluate the learning efficacy of the *Meta*-CNN algorithm across different CNS drug categories related to distinct phenotypes, we individually examined learning curves for each category, ranging from ATC codes N01 (anesthetics) to N07 (other nervous system drugs) (Figure S7). The results suggest that the *Meta*-CNN algorithm exhibited enhanced stability and predictive accuracy than the CNN-only model (Figure S8). For drugs in all categories spanning from N01 to N07, the *Meta*-CNN model all demonstrated some degree of improvement in testing accuracy. Particularly for the N04: anti-Parkinson category, we achieved a 72.5% accurate rate after 191 iterations. There were another two drug categories surpassing the average performance, including N01: anesthetics and N02: analgesics, with the best accuracy scores being 67.0% and 57.0%, respectively (Figure 3F). These results suggest that, in the context of identifying the coherent correlation between the BAM clusters and ATC clinical drug categories, a meta-learning component significantly improves the feature extraction capabilities, which would be greatly beneficial for subsequent pharmacological studies.

A focus on anti-Parkinson drugs (ATC codes: N04)

A prevalent neurodegenerative disorder, second only to Alzheimer's disease, Parkinson's disease (PD) represents a substantial challenge to the healthcare systems and socioeconomic structures globally and is exacerbated by an increasingly aging population.[49] Current drug treatments for PD, listed under the N04: anti-Parkinson category in the ATC classification, are mainly palliative in nature, aimed at mitigating the symptomatic progression of the disease. These pharmacological interventions predominantly include dopamine agonists (N04B), which function by emulating the neurochemical actions of dopamine within the brain.[50] As the scientific community strives for an in-depth comprehension of the pathophysiology underlying PD, the development of more potent, disease-modifying therapeutic strategies is a focal point of much ongoing research. In our analysis of the N04 (anti-Parkinson) class of drugs using the high-throughput BAMing strategy, it was noted that the pharmacological targets were primarily dopamine receptors with mostly agonists, which have been verified as efficacious in PD management (Figures 4A and 4B, also see Figure S9). Except for these main pharmacological targets, during our analysis of protein target pathways, it was observed that numerous drugs within the N04 (anti-Parkinson) class category exhibit a complex hierarchy of protein-specific targets. This intricate structure could also serve as a mechanism for verifying potential N04 (anti-Parkinson) category drugs. (Figure 4C, more details in Figure S9C).

While the BAMs acquired from these drugs exhibited some common patterns, which generated different clusters (Figure 4D), distinct variations were also discernible suggesting the possibility of extracting patterns of induced functional activity for different drugs. Besides, the learning curve results illuminate interesting aspects of the N04 drug category. Using a CNN-only network, the average predictive accuracy for N04 (anti-Parkinson) reached its maximum point at 65.4% at the 29th iteration (n = 9) (Figure S8), followed by an inconsistent decline. However, with the integration of the meta-learning algorithm into the system, the learning curve displayed a consistent and gradual increase over iterations, with a peak predictive accuracy of 72.5% at the 191st iteration (Figure 3F). This observation implies that the meta-learning algorithm enhances the stability of the pharmacological prediction system and is particularly useful when evaluating BAMs of anti-Parkinson drugs, which leads us to further develop a predictive system for future screening of N04 drug leads using the BAMs as an input.

Theoretical analysis of the two-phase meta-learning network

Meta-learning, featured as "learning to learn", revolves around the idea of developing models that can learn new tasks expeditiously, despite only being trained by a limited number of examples.[51] The essence of meta-learning is to harness the knowledge garnered from past tasks,

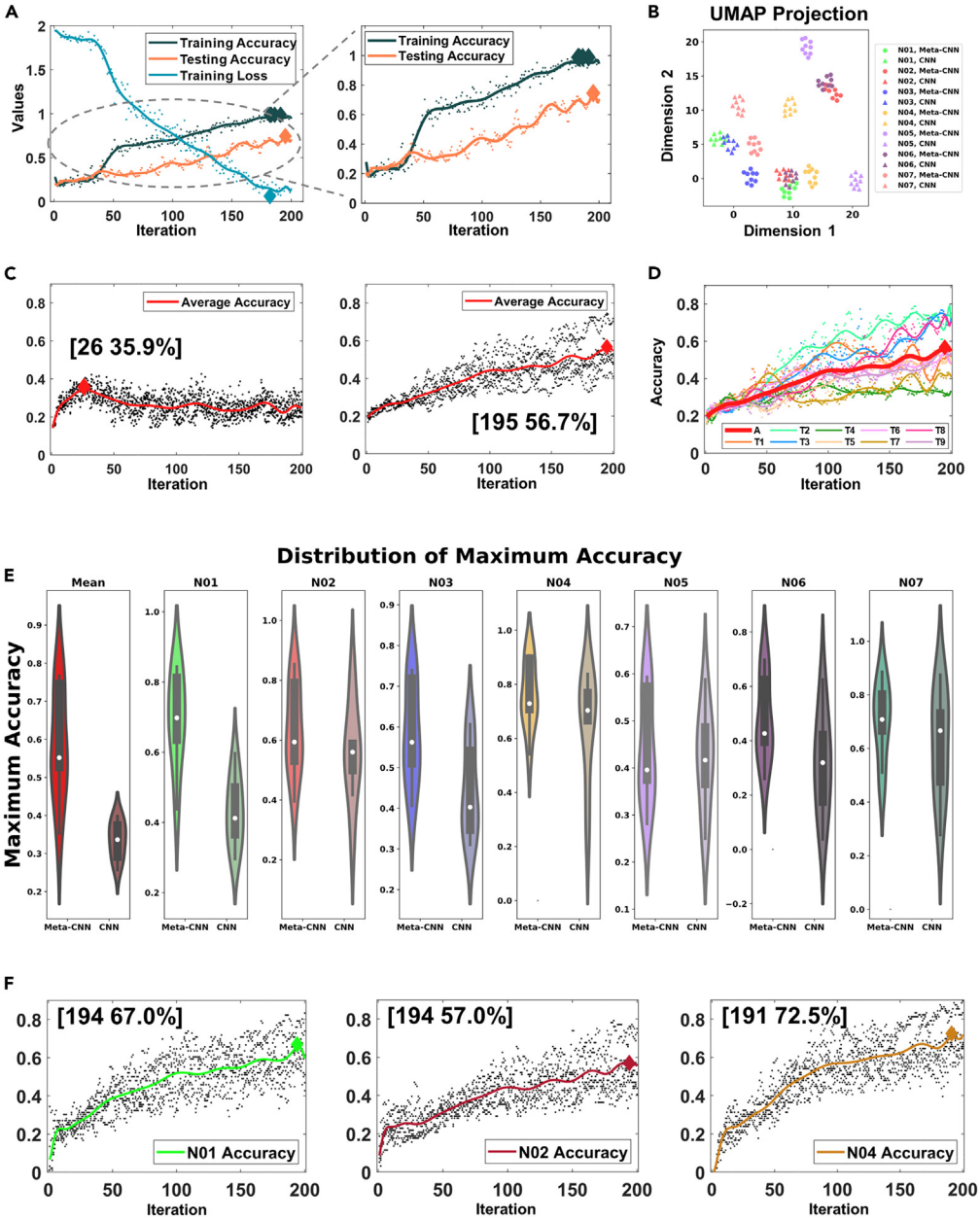


Figure 3. Results for ATC-BAM experiments

(A) Illustration of the accuracy and loss results from a Meta-CNN experiment.

(B) UMAP projection outcomes for two configurations of network models (final 50 iterations used, $n = 9$).

(C) The figure delineates the classification accuracy across a range of experiments conducted with diverse random distributions of dataset. From left to right: the testing accuracy of the CNN and Meta-CNN learning models (red line: average accuracy, black dots: individual experiment accuracy values for each iteration, with outlier data points being omitted. The peak accuracy achieved between the 20th and 200th iterations is accentuated with a diamond symbol).

(D) Variation curves of $n = 9$ experiments for the Meta-CNN learning network (red line: average accuracy, T1-9 colored lines: accuracy curves of each experiment).

(E) For 180th to 200th iterations, the spread of the maximum accuracy values observed across nine trials ($n = 9$) under Meta-CNN and CNN-only models. Data are represented as a violin plot showing the distribution, median, and IQR.

(F) The average accuracy trajectories for the drug categories ATC codes N01, N02, and N04 during the testing phase.

employing this wisdom to facilitate and expedite the learning process for new tasks. This is especially useful when the available training data for the new task is sparse or limited. Meta-learning thereby aims to design machine learning models that can generalize from a few examples, a concept known as few-shot learning.[52,53] In our study involving high-throughput whole-brain activity mapping, the dataset, though

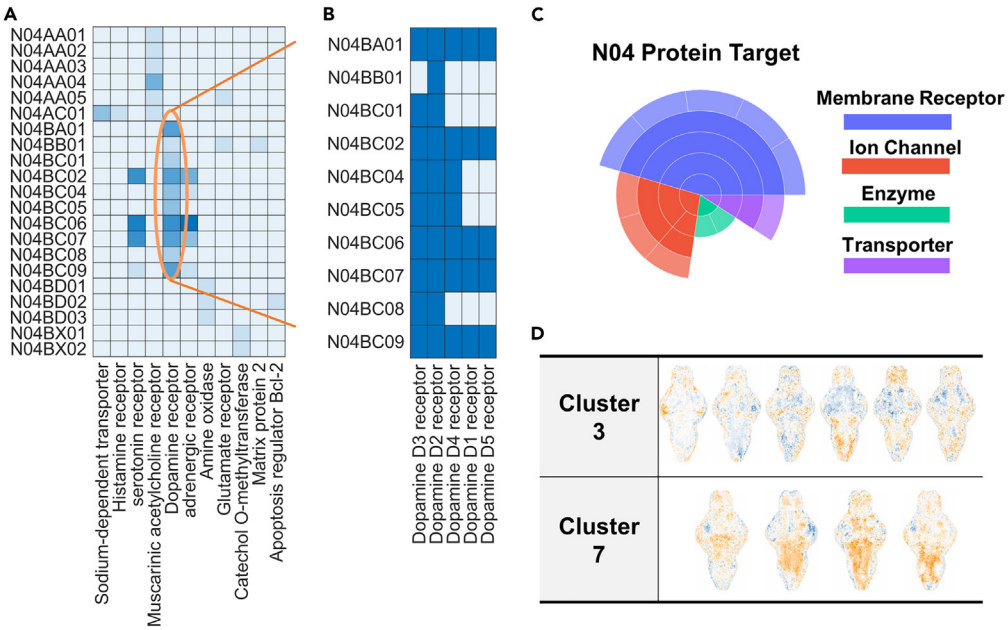


Figure 4. A comprehensive overview of biomedical targets and BAM visuals related to N04 category anti-Parkinson pharmaceutical agents

(A) Biological target active heatmap of drugs in N04 ATC category.

(B) Functional assemblies within the dopamine receptor subcategory.

(C) Different hierarchy of protein-oriented targets specific to N04 pharmaceutical solutions.

(D) BAM images correlating with the same clusters.

extensive, is still not sufficiently large to train robust traditional machine learning models, despite our efforts to map more than half of the CNS drugs registered with whom (Table S1). To address these data scarcity problem, we used meta-learning for constructing few-shot validation models, even amid complex biological systems with limited data availability.[23] We leveraged shared knowledge from one domain to aid in tasks related to another domain. This strategy led to a model with significantly enhanced learning capability and adaptability. In this study, we extended our previous research on bilevel optimization,[37,54–56] which has demonstrated enhanced efficacy in addressing multi-layered problems. Meanwhile, meta-learning, characterized by its hierarchical structure comprising meta-training and meta-testing layers, aligns well with the principles of bilevel optimization.[37] Our objective was to amalgamate the processes of hyperparameter training and meta-training from a meta-learning perspective, facilitating the simultaneous acquisition of training variables and hyperparameters.[54,57] This synergy renders bi-level optimization a particularly apt approach for our study. We proposed to frame the meta-learning task within the bilevel optimization framework, which also featured a hierarchical structure. In bilevel optimization, the upper-level optimization seeks the optimal hyperparameter combination, while the lower-level optimization focuses on minimizing the model's loss function for a given hyperparameter set (Figure 5A). This bi-level meta-learning approach incorporates knowledge shared between tasks from a meta-training domain $D_{m-train}$ into a meta-testing domain by utilizing a pre-trained network at D_{m-test}, where the cross-task intermediate representation layers are adapted to train the base model on a previously unseen task at D_{m-test} (Figures 5B and 5C). In this few-shot learning network, the BAM datasets were labeled by corresponding ATC code: N01 to N07. In the meta-training phase, a handful of BAM data sufficed to generate a robust set of hyper-parameters encapsulating some general features. These hyper-parameters in the meta-testing phase empowered the network to generate more nuanced features for BAMs spanning different drug categories, even though the data size was considered to be limited for traditional machine learning (Figures 5B and 5C). A crucial deviation from our prior validation methods using functional clustering approaches is that, in this system, we calculated the validation accuracy directly using a pharmacologically labeled dataset during the testing phase. This approach provided a direct evaluation of the learning efficiency of our model. For drug discovery, meta-learning has shown its promise in predicting the properties of chemical compounds based on sparse examples of related molecules.[58] The dearth of extensive, high-quality experimental molecular data makes meta-learning methodologies particularly valuable for drug discovery tasks, providing an effective solution to conquer such challenges. In fact, this approach has shown great potential to improve the drug discovery process by making it more efficient and cost-effective.

BAM and CDA for different pharmacological discovery

To establish a reference for the BAM model, we then tested the *Meta*-CNN model for the chemical descriptor for the same drug/compound library that was used in our high-throughput BAMing test.[9] As the chemical structure of compounds is related to the physicochemical properties, furthermore determines its absorption, distribution, metabolism, excretion, and toxicity (ADME/Tox) properties, and synthetically

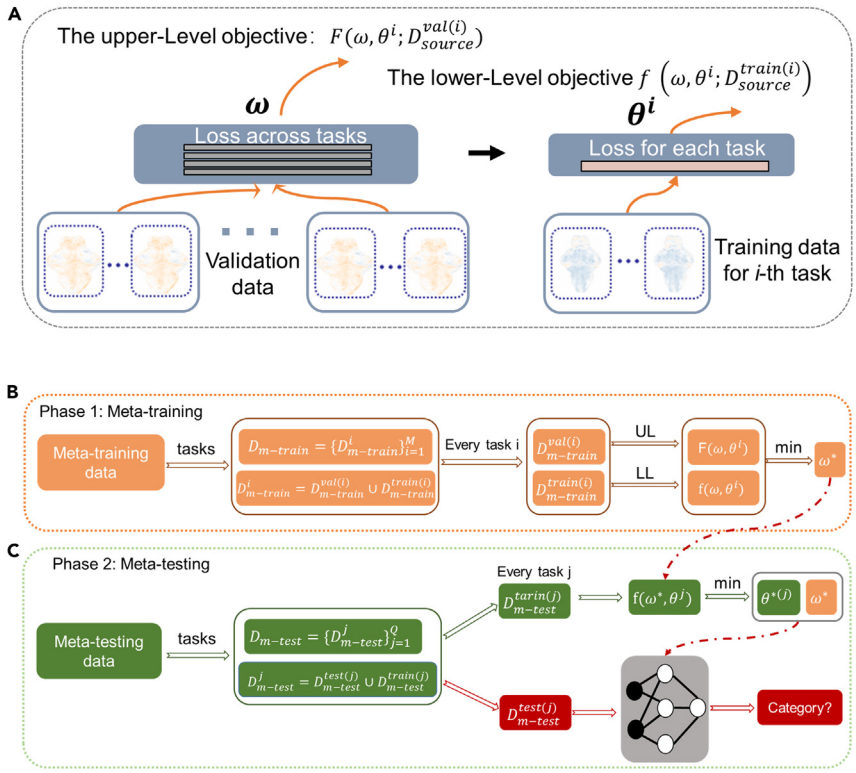


Figure 5. Flow diagram of a two-phase meta-learning algorithm

(A) Structure of bi-level optimization, the upper-level optimization seeks the optimal hyperparameter combination, while the lower-level optimization focuses on minimizing the model's loss function for a given hyperparameter set.

(B) Meta-training phase for calculating primary hyper-parameter shared between the tasks.

(C) Meta-testing phase is critical for the adjustment of the optimal network parameters and the computation of the accuracy for each task, leading to the classification of the final category.

affects the pharmacological property. With the high efficiency and accessibility in data acquisition, chemical molecular properties and fingerprint information, based on drug-likeness, provide potential for pre-clinical high throughput screening of CNS drug discovery.[59,60] We further established a library that incorporated a chemical descriptors array (CDA) (Figure 6A), which was generated from three-dimensional constitutional, topological, and pharmacophore data. Prior to experiment with the *meta*-CNN, an unsupervised learning protocol was employed to cluster the CDA library (STAR Methods detailed in Figure S1). Analysis revealed that selecting ten classes optimally shaped the cumulative distribution function (CDF) curves. Moreover, the consensus matrix, affirmed effective data clustering within the CDA library (Figure S10). The clustering results indicated a notable association: compounds in the N01 ATC category predominantly clustered within CDA class 3, as evidenced by a hypergeometric test value of 0.04. Similarly, compounds in the N03 ATC category showed a significant association with CDA class 2, marked by a hypergeometric test value of 0.03 (Figures S10 and S11). In the CNN-only learning experiments (n = 9), the best average testing accuracy improved from 14.3% to 24.7%, marking a 10.4% increase. Using *meta*-CNN, the best average testing accuracy improved to 27.1%, reflecting a 12.8% increase (Table 3). Both the CNN-only and the *Meta*-CNN displayed a gradual and consistent increase in average testing accuracy across iterations when compared with the BAM dataset (Figure 6B, detailed in Figures S12 and S13). For individual drug category, the learning capacity of the meta-learning algorithm only showed effective improvements for N01: anesthetics and N04: anti-Parkinson drugs (Figure 6C). Compared to the enhancement achieved by the *Meta*-CNN model in the learning process of the BAM library (Figure 3), the improvement in the CDA library was less pronounced. These results suggested that the meta-learning algorithm was substantially more effective in improving the analysis of the BAM dataset, which is richer and more complex for information extraction. In the analysis of learning curves under the *Meta*-CNN network, the prediction accuracy of the BAM library was consistently superior to that of the CDA library across all iterations (Table 4). This supremacy was further evidenced by the analysis of applying the *Meta*-CNN model to different datasets (n = 9), we included a statistical comparison table summarizing the results of paired t tests, Wilcoxon signed-rank tests, and Cohen's d effect sizes for the performance of the *Meta*-CNN model based on CDA and BAM applications across two metrics: final accuracy and maximum accuracy (Figure 6D, detailed in Table S6). The results demonstrate the superior performance of the *Meta*-CNN model over the CNN model on the BAM dataset in terms of both final and maximum accuracy. Additionally, the BAM dataset performs better than CDA using these models. The statistical tests and effect size calculations confirm that these differences are not only statistically significant but also practically

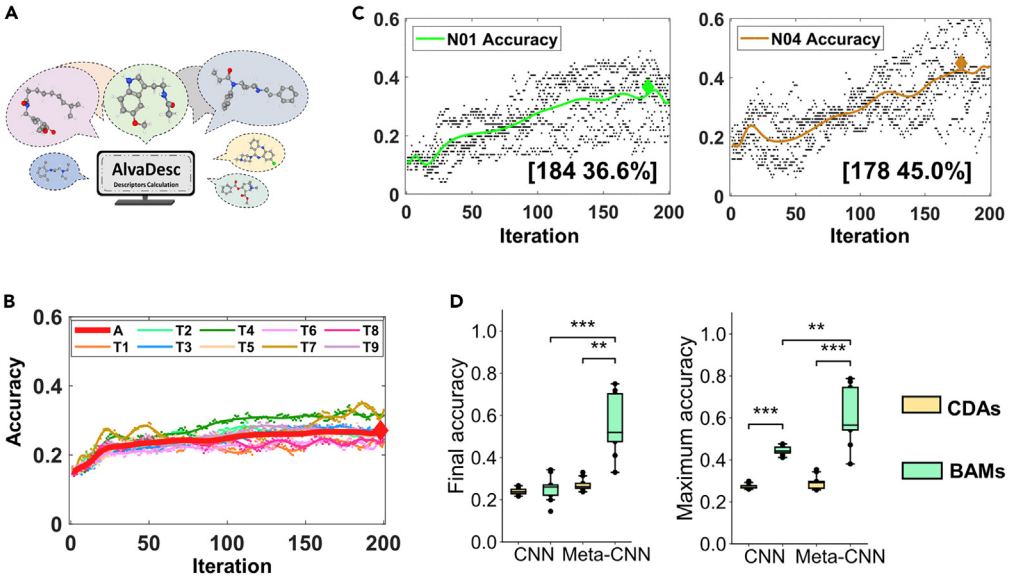


Figure 6. Results for ATC-CDA experiments

(A) Diagrammatic representation of the chemical descriptors array (CDA).

(B) Variation curves of $n = 9$ experiments for the Meta-CNN learning network (red line: average accuracy, T1-9 colored lines: accuracy curves of each experiment).

(C) The average accuracy trajectories for the drug categories N01, and N04 during the testing phase.

(D) The boxplots of final accuracy and maximum accuracy distribution. Data are represented as boxplots, including median and interquartile range (IQR), with whiskers extending to 1.5 times the IQR, ** indicates $p < 0.01$, and *** indicates $p < 0.001$ by t test.

meaningful. This highlights the robustness and effectiveness of the *Meta*-CNN model in comparison to other control models, as well as its ability to capture features in BAM images of drug different categories.

Validation and drug repositioning results

Pharmaceutical development, particularly in the realm of CNS therapies, is a complex and time-consuming endeavor. Traditional drug development, often reliant on *de novo* design, involves significant time and resources before market introduction. In recent years, there has been a shift toward drug repositioning, a strategy aiming to repurpose the therapeutic applications for existing drugs, thereby extending their utility beyond initial indications.[61,62] Our study integrated the *Meta*-CNN model into drug repositioning, focusing on CNS therapeutic classification. To validate the model's effectiveness in this context, we conducted comprehensive testing using an untrained, CNS-specific BAM dataset (more details in Table S7). This was crucial to assess the model's adaptability to drug repositioning challenges. Subsequently, we explored the potential of repurposing non-N (nervous system) series ATC classified drugs for CNS therapies (Figure 7A, more details in Table S8). From the verification results, we found that the *Meta*-CNN model demonstrated high classification accuracy in the ATC: N03 (anti-epileptics) to N06 (psychoanaleptics) categories. Notably, in the N04 (anti-Parkinson) drug category, mean, and median values were 55.8% and 60%, respectively, surpassing the critical threshold of 0.5 (Figure 7B). This indicates the *Meta*-CNN model's robust discriminative capability. Further analysis revealed the model's precision, with five compounds notably exceeding the 50% classification threshold (Figures 7C and 7D). This was particularly evident for the compound dextromethorphan (R05DA09), an antitussive medication, which displayed a significant associative value of 0.71 in the matrix, indicative of a strong predictive correlation within the N04 (anti-Parkinson) category. The model's alignment with BAM cluster 7 and the N04 (anti-Parkinson) biological target map further validated the algorithm's acuity in identifying compatible target profiles within the N04 (anti-Parkinson) drug category (Figures 4D and 7E, detailed in Figure S9). This was complemented 'by evidence positioning dextromethorphan as a promising candidate in PD applications, aligning with contemporary research findings.[63] Furthermore, our analysis included sinecatechins (D06BB12), a water-soluble extract derived from green tea (*Camellia sinensis*) leaves, which exhibited a matrix value of 0.41. Although no direct evidence links sinecatechins to the treatment of PD, their classification under catechins, a group receiving attention in neurology and neurodegeneration studies, is significant. Research in these areas has focused on key molecular mechanisms like oxidative stress, inflammation, and apoptosis, all relevant to conditions like Parkinson's and Alzheimer's disease.[64] In summary, the *Meta*-

Table 3. 2 network configurations for optimal testing accuracy: CNN model ($n = 9$), and *Meta*-CNN learning algorithm ($n = 9$)

Input	Label	Algorithm	Training	Validation	Training	Testing	Average accuracy
CDAs	ATC	CNN	18-shot	–	–	3-shot	24.7%
CDAs	ATC	*Meta*-CNN	4-shot	3-shot	11-shot	3-shot	27.1%

Table 4. Comparison between *Meta*-CNN based experiments on CDA and BAM datasets

Meta-CNN comparison	–	t test *p* value	Wilcoxon *p*-value	Cohen's d Effect Size
CDA vs. BAM	Final accuracy	1.03E-03	3.91E-03	−2.61
CDA vs. BAM	Maximum accuracy	4.38E-04	3.91E-03	−3.05

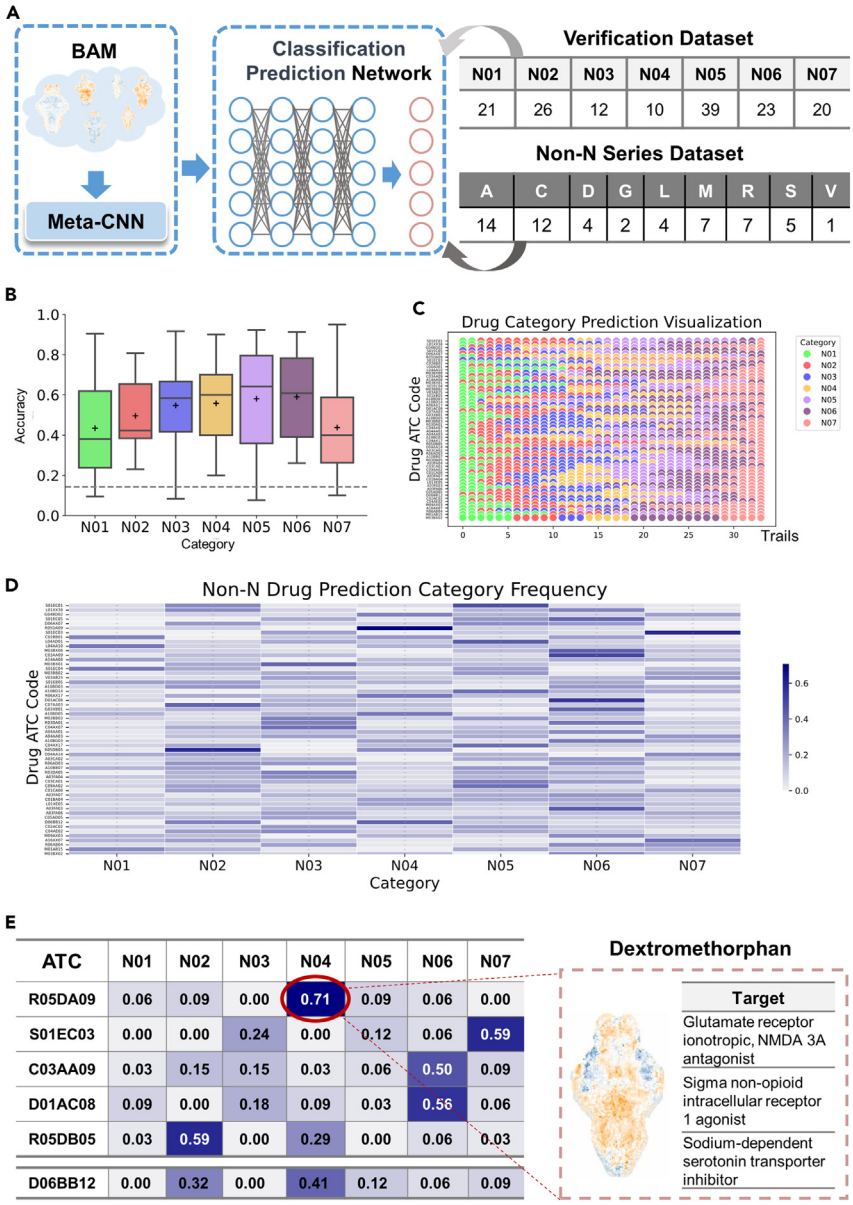
The statistical analysis results of the t test, Wilcoxon signed-rank test, and Cohen's d effect size are included.

CNN model demonstrates effectiveness in CNS drug classification and suggests a more efficient approach to CNS drug discovery, particularly for the N04 (anti-Parkinson) category. This highlights the potential of a data-centric approach in leveraging artificial intelligence for systems neuropharmacology and pharmaceutical innovation.

DISCUSSION AND CONCLUSION

With the ongoing and effortless progress in developing therapies for CNS disorders, there is an increasing need for effective animal models and analytical methods to facilitate the discovery of more compounds with potential clinical usage.[65] Streamlining research timelines and reducing costs is crucial in enhancing compound identification and decision-making processes during the initial stages of therapeutic development, ultimately facilitating the discovery of innovative and cost-effective CNS therapies.[9] In pharmacological research and development, the integration of machine learning, notably deep neural networks, has markedly advanced the field of drug development. These models demonstrate exceptional predictive capabilities in determining the properties and activities of small molecule compounds.[11] However, traditional machine learning algorithms often encounter challenges in CNS drug screening due to the limited number of approved drugs, scarce training samples, and overall large search space. To address these challenges, meta-learning-based, few-shot algorithms offer a solution by leveraging prior experience to tackle fresh problems.[23] This reduces the time and cost associated with testing biological samples, thereby accelerating the establishment and application of predictive models.

This study focused on the potential of integrating high-throughput brain activity mapping[24] with advanced machine learning methods, particularly few-shot meta-learning algorithms, to address the mentioned challenges in CNS drug discovery. Our previous research has introduced explanations and analyses regarding patterned features in the brain,[24] demonstrating that common features of BAMs could be linked to specific effects on CNS physiology shared among drugs within the same cluster. This innovative use of BAM data allows us to bypass traditional reliance on chemical structure or single molecular targets, instead generating drug assessments purely based on physiological phenotypes. In Figure S1C, it is clear that except for category N05, all other categories exhibit a relatively small *p* value in evaluation of association with corresponding clusters in the hypergeometric test. This consistency aligns with the classification results obtained from the *Meta*-CNN model for the classification task (Figure S7). This observation implies that the classification of CNS drugs is associated with the similarity or corresponding features among patterned BAMs. The category with better feature capture properties in unsupervised learning tends to yield better predictive results in our *Meta*-CNN model. Compared with other commonly used drug discovery strategies, such as recombinant DNA technology and tissue culture techniques, the early stages of pharmaceutical discovery often involve the use of recombinant proteins or heterologous cellular models.[66] These methods allow for the detailed study of specific biological targets and the screening of compounds in a controlled environment. More recently, screening methods using organism models have been developed.[26] These platforms offer the advantage of evaluating drug effects in the context of whole-organism biology, providing insights into organ-specific physiology and potential systemic effects. However, these methods are limited in their ability to perform large-scale chemical screens. The complexity and variability inherent in whole-organism models present significant challenges for high-throughput screening, which is essential for the rapid identification of potential therapeutic compounds. In our approach, we utilized HT-BAMing technology, which enables rapid assessment of changes in brain physiology and activity in response to compound exposure at the cellular resolution level across the entire zebrafish brain. BAM data serve as evaluation inputs, providing the potential to directly correlate zebrafish brain physiology with the clinical potency of compounds. This innovative use of BAM data allows us to bypass traditional reliance on chemical structure or single molecular targets, generating drug assessments based purely on physiological phenotypes. This provides a more accurate and physiologically relevant assessment of drug effects and is the rationale behind choosing BAM as our dataset for analyzing the drug discovery few-shot learning model. Essentially, BAM captures intricate functional details of the brain at a single cell level, providing a rich source of phenotypic data that reflects the complexity and dynamics of complex and dynamic biological systems. However, the high complexity and scarcity of BAM data also pose challenges for machine learning-based analytical strategies, especially, since traditional machine learning algorithms were limited in generating robust and generalizable models for effective pharmacological prediction. Few-shot meta-learning algorithms, also known as "learning to learn" methods, leverage the concept of transfer learning, using knowledge from prior tasks to facilitate the learning of new ones. By employing few-shot meta-learning, our approach aims to exploit its high adaptability and data efficiency, making it particularly suitable for CNS drug discovery applications where new tasks frequently arise. This approach addresses several limitations inherent in transfer learning and reinforcement learning, especially in scenarios characterized by limited data availability and the necessity for swift adaptation. In our study, we explored the integration of the few-shot *Meta*-CNN model and observed significant improvements in model performance for the prediction of the compound's clinical usage (Figure 2). By leveraging the inherent ability of few-shot meta-learning algorithms to draw on prior knowledge and adapt swiftly to new tasks, we could effectively extract indicative patterns from the BAM data using only a handful of



ATC	N01	N02	N03	N04	N05	N06	N07
R05DA09	0.06	0.09	0.00	0.71	0.09	0.06	0.00
S01EC03	0.00	0.00	0.24	0.00	0.12	0.06	0.59
C03AA09	0.03	0.15	0.15	0.03	0.06	0.50	0.09
D01AC08	0.09	0.00	0.18	0.09	0.03	0.58	0.06
R05DB05	0.03	0.59	0.00	0.29	0.00	0.06	0.03
D06BB12	0.00	0.32	0.00	0.41	0.12	0.06	0.09

Figure 7. Results of validation testing and pharmaceutical repositioning

(A) Experimental methodology involving two processes: one exploits a segment of untrained N series BAM data for verifying the operational capacity of the central nervous system (CNS) drug classification system derived from both the BAM library and our Meta-CNN model, while the other uses a non-N series drug repository for forecasting potential CNS drugs from this series.

(B) A comparative boxplot of the validation proceedings ($n = 9$, random 5 values from iteration 180^{th}–200^{th}). Remarkably, from ATC categories N04 through N06, all exhibit a high successful classification rate, with both average and median values exceeding 0.5. Data are represented as boxplots, including median and interquartile range (IQR), with whiskers extending to 1.5 times the IQR.

(C and D) The drug repositioning results for the non-N series dataset ($n = 9$, random 5 results were chosen from iteration 180^{th}–200^{th}, with overfitting data results being omitted).

(E) On the left, it displays classification frequencies surpassing 50% resulting from drug repositioning trials. On the right, it highlights a singular outcome within the N04 category where the drug dextromethorphan exhibits similar biological targets and BAM imagery to other drugs within the same category.

examples. This promises a significant acceleration in the process of model training and evaluation, thus reducing the time and resources required for the prediction of drug-related CNS properties. Our results indicate that the proposed method achieves a high accuracy in predicting brain activity changes and clustering phenotypic responses to drug treatments. This demonstrates the robustness and reliability of our approach in identifying meaningful patterns in neuropharmacological data. Notably, our method outperforms traditional machine learning

algorithms, such as random forest, highlighting its effectiveness in handling small sample sizes and complex data structures prevalent in neuropharmacology (Figures 2 and 3), Moreover, we found that the meta-learning-based CNN model exhibited better stability and higher predictive performance, especially for N04 (anti-Parkinson) category, with a stable increase from 14.3% to 72.5%. (Figures 3, 4, and 6). Furthermore, the drug repositioning results reveal potential uses for existing drugs. This capability can significantly accelerate the drug discovery process by identifying more therapeutic applications for known compounds, thereby potentially reducing the time and costs for the development of therapeutics targeting complex brain disorders. We validated our predictions by comparing the results with known pharmacological data. The consistency between our predictions and experimental outcomes supports the validity of our approach (Figure 7).

Still, it is recognized that there could be hurdles to translate findings in zebrafish models to human clinical applications. Although zebrafish are a well-established model for studying vertebrate development and disease, some differences between zebrafish and human biology must be taken into account, especially for clinical applications,[6] this is exactly why we employed the ATC system to supervise the AI learning of the BAM data, which is meant to directly analyze and predict the potential of clinical usage for a chemical compound. The only information required is the induced BAM generated in larval zebrafish. In this sense, our strategy provides a paradigm of using the zebrafish for drug discovery and bypasses the problem induced by the cross-species biological differences by an independence of chemical or molecular information. Clearly, drug development or repositioning necessitates rigorous clinical trials to ensure safety and efficacy in humans.[2] As one potential caveat, variability in brain activity data due to the biological differences among individual larvae could impact the analytical results.[9] While we here only differentiate 2D brain regions in a z-projected image, and use the associated BAM information for computational analysis, as an extension of this screening paradigm, 3D brain structural information can be included for analysis with the help of more advanced imaging techniques. In summary, our approach shows significant promise in drug repositioning and neuropharmacology study, further studies need to be done for clinical translation. Other drug discovery methods, such as structure-based drug design, CDA datasets, and proteomics and genomics studies, offer valuable insights into drug action, toxicity, resistance, and efficacy,[67] making them indispensable in the drug discovery process. Bioinformatics and computational approaches hold promise in enhancing our ability to integrate genomics, proteomics, and 3D-structural biology data to refine the subset of proteins likely modulated by a drug.[68] These datasets could provide diverse patterns for generating categorized CNS drug features. By collecting these multimodal datasets, we could potentially improve the predictive ability of our model. In future studies, it would be interesting to investigate the application of multi-modal data integration combined with different learning models to further enhance our understanding and predictive capabilities in drug discovery and neuropharmacology. By improving data quality, optimizing the model, and conducting cross-feature validation, we believe that the robustness and clinical relevance of our findings could be well improved in future studies. (Figure 7).

The few-shot learning framework addresses concerns regarding the translation of experimental datasets to clinical settings, which are characterized by diverse contexts and limited data. This framework provides a bridge from the extensive sample data gathered in high-throughput screens (n-of-many) to the distinctive contexts of individual patients (n-of-one).[23] Personalized medicine incorporates an individual's genomic, environmental, and lifestyle data to guide medical decisions.[69] In this research, it has been demonstrated that common features of BAMs can be linked to specific effects on CNS physiology, the integration of few-shot meta-learning algorithms with clinical images holds immense potential for both personalized medicine and precision drug discovery. By leveraging the adaptive nature of meta-learning to glean insights from limited data, our approach could be finely tuned to individual patient profiles, enabling more precise and effective therapeutic interventions. This entails curating personalized clinical image profiles for patients, facilitating the identification of the most suitable CNS drugs based on their distinct brain activity patterns. Furthermore, the rapid and accurate classification of drugs using image datasets opens avenues for repurposing and repositioning existing pharmaceuticals, potentially reducing the time and cost associated with drug development. Future research will focus on expanding our dataset, integrating multi-omics data, and enhancing model interpretability. Such efforts aim to develop a robust, patient-specific drug discovery platform that enhances the efficacy and safety of CNS therapeutics.

Limitations of the study

In this study, we used zebrafish with the elval3:GCaMP5G line and conducted experiments on zebrafish larvae aged 6–8 days post-fertilization without distinguishing the sex of the experimental vertebrates. The variation in the brain activity maps among different biological replicates poses some uncertainty in the feature extraction and machine learning for the identification of BAM-associated neuropharmacological phenotypes. Additionally, while zebrafish models are well-established for studying vertebrate development and disease, the inherent biological differences between zebrafish and human can possibly affect the translation of findings from this study to human clinical applications.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Peng SHI (pengshi@cityu.edu.hk).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table.

- The data have been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to analyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (U20A20194, 2222106, 12326605, 62331014), by General Research Fund (11215920, 11220024, 11218522, 11218523) from the Research Grants Council of Hong Kong SAR, Guangdong Basic and Applied Basic Research Foundation (2022B1515020082), Shenzhen-Hong Kong-Macau Science and Technology Program (Category C, SGDX20201103093005O2), and Shenzhen Science and Technology Program (RCYX20200714114700072). Support from Innovation and Technology Commission of Hong Kong through the Centre for Cerebro-Cardiovascular Health Engineering and funds from City University of Hong Kong (7005084, 7005206, 7005642, 7020003, 7020077, 9680233, 9240060) are also acknowledged.

AUTHOR CONTRIBUTIONS

P.S. and J.Z. conceived and supervised the project. X.L., Y.D., X.W., J.Z., and P.S. designed the research. X.L. performed the high-throughput screening and analyzed the data. J.Z. developed the bilevel few shot learning framework. Y.D. and Y.C. implemented the computations and performed the informatic analysis. X.L., Z.L., and S.J.H. constructed the chemical library. S.H.C. helped with zebrafish maintenance. S.Z. contributed to data processing. X.L., Y.D., and P.S. wrote the manuscript. Y.C., W.Z., H.L., S.J.H., and J.Z. also contributed to the writing of the manuscript.

DECLARATION OF INTERESTS

P.S. is listed as an inventor on a patent (US 9,897,593) filed by the City University of Hong Kong describing the system for automated handling of larval zebrafish. S.J.H. has served or serves on the advisory board of Proximity Therapeutics, Psy Therapeutics, Frequency Therapeutics, Souvien Therapeutics, Sensorium Therapeutics, 4M Therapeutics, Ilios Therapeutics, Entheos Labs, the Alzheimer's Disease Drug Discovery Foundation, and the Kissick Family Foundation FTD grant program, none of whom were involved in the present study. S.J.H. has also received speaking or consulting fees from Amgen, AstraZeneca, Biogen, Merck, Regenacy Pharmaceuticals, Syros Pharmaceuticals, and Juvenescence Life, as well as sponsored research or gift funding from AstraZeneca, JW Pharmaceuticals, Lexicon Pharmaceuticals, Vesigen Therapeutics, Compass Pathways, Atai Life Sciences, and Stealth Biotherapeutics. The funders had no role in the design or content of this article or the decision to submit this review for publication.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Zebrafish line
- METHOD DETAILS
  - Microfluidic immobilization and drug exposure
  - Image acquisition and T-score BAM generation
  - Molecular descriptor database
  - The data pre-processing workflow
  - Data partition
  - Meta-learning algorithm design
  - The neural network models
  - Code package
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analyses and hypothesis testing procedures

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110875.

Received: March 11, 2024
Revised: June 10, 2024
Accepted: August 30, 2024
Published: September 3, 2024

REFERENCES

1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet *392*, 1789–1858.
2. Gribkoff, V.K., and Kaczmarek, L.K. (2016). The Need for New Approaches in CNS Drug Discovery: Why Drugs Have Failed, and What Can Be Done to Improve Outcomes. Neuropharmacology *1*, S0028390816300934.
3. Nance, E., Pun, S.H., Saigal, R., and Sellers, D.L. (2022). Drug delivery to the central nervous system. Nat. Rev. Mater. *7*, 314–331.
4. Abbott, N.J., Patabendige, A.A.K., Dolman, D.E.M., Yusof, S.R., and Begley, D.J. (2010). Structure and function of the blood-brain barrier. Neurobiol. Dis. *37*, 13–25.
5. Bors, L.A., and Erdő, F. (2019). Overcoming the blood–brain barrier. challenges and tricks for CNS drug delivery. Sci. Pharm. *87*, 6.
6. Khan, K.M., Collier, A.D., Meshalkina, D.A., Kysil, E.V., Khatsko, S.L., Kolesnikova, T., Morzherin, Y.Y., Warnick, J.E., Kalueff, A.V., and Echevarria, D.J. (2017). Zebrafish models in neuropsychopharmacology and CNS drug discovery. Br. J. Pharmacol. *174*, 1925–1944.
7. Balaban, A.T. (2009). Drug Design, Molecular Descriptors. In Encyclopedia of Complexity and Systems Science, R.A. Meyers, ed. (Springer New York), pp. 1–31.
8. Barnash, K.D., James, L.I., and Frye, S.V. (2017). Target class drug discovery. Nat. Chem. Biol. *13*, 1053–1056.
9. Sun, D., Gao, W., Hu, H., and Zhou, S. (2022). Why 90% of clinical drug development fails and how to improve it? Acta Pharm. Sin. B *12*, 3049–3062.

10. Leung, M.K.K., Delong, A., Alipanahi, B., and Frey, B.J. (2016). Machine learning in genomic medicine: a review of computational problems and data sets. Proc. IEEE *104*, 176–197.
11. Yip, K.Y., Cheng, C., and Gerstein, M. (2013). Machine learning and genome annotation: a match meant to be? Genome Biol. *14*, 205.
12. Alipanahi, B., Delong, A., Weirauch, M.T., and Frey, B.J. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nat. Biotechnol. *33*, 831–838.
13. Ballester, P.J., and Mitchell, J.B.O. (2010). A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. Bioinformatics *26*, 1169–1175.
14. Califano, A., and Alvarez, M.J. (2017). The recurrent architecture of tumour initiation, progression and drug sensitivity. Nat. Rev. Cancer *17*, 116–130.
15. Carro, M.S., Lim, W.K., Alvarez, M.J., Bollo, R.J., Zhao, X., Snyder, E.Y., Sulman, E.P., Anne, S.L., Doetsch, F., Colman, H., and Lasorella, A. (2010). The transcriptional network for mesenchymal transformation of brain tumours. Nature *463*, 318–325.
16. Marcus, G. (2018). Deep Learning: A Critical Appraisal. Preprint at arXiv. https://doi.org/10.48550/arXiv.1801.00631.
17. Dong, Z., Zhang, R., Shao, X., and Kuang, Z. (2020). Learning sparse features with lightweight ScatterNet for small sample training. Knowl. Base Syst. *205*, 106315.
18. Argyriou, A., Evgeniou, T., and Pontil, M. (2006). Multi-task feature learning. Adv. Neural Inf. Process. Syst. *19*, 1.
19. Blitzer, J., McDonald, R., and Pereira, F. (2006). Domain Adaptation with Structural Correspondence Learning (Association for Computational Linguistics), pp. 120–128.
20. Vrbančič, G., and Podgorelec, V. (2020). Transfer Learning With Adaptive Fine-Tuning. IEEE Access *8*, 196197–196211.
21. Arulkumaran, K., Deisenroth, M.P., Brundage, M., and Bharath, A.A. (2017). Deep Reinforcement Learning: A Brief Survey. IEEE Signal Process. Mag. *34*, 26–38.
22. Dulac-Arnold, G., Levine, N., Mankowitz, D.J., Li, J., Paduraru, C., Gowal, S., and Hester, T. (2021). Challenges of real-world reinforcement learning: definitions, benchmarks and analysis. Mach. Learn. *110*, 2419–2468.
23. Ma, J., Fong, S.H., Luo, Y., Bakkenist, C.J., Shen, J.P., Mourragui, S., Wessels, L.F.A., Hafner, M., Sharan, R., Peng, J., and Ideker, T. (2021). Few-shot learning creates predictive models of drug response that translate from high-throughput screens to individual patients. Nat. Cancer *2*, 233–244.
24. Lin, X., Duan, X., Jacobs, C., Ullmann, J., Chan, C.Y., Chen, S., Cheng, S.H., Zhao, W.N., Poduri, A., Wang, X., et al. (2018). High-throughput brain activity mapping and machine learning as a foundation for systems neuropharmacology. Nat. Commun. *9*, 5142.
25. Chiang, M., Back, H.M., Lee, J.B., Oh, S., Guo, T., Girgis, S., Park, C., Haroutounian, S., and Kagan, L. (2020). Pharmacokinetic modeling of the impact of P-glycoprotein on ondansetron disposition in the central nervous system. Pharm. Res. (N. Y.) *37*, 205.
26. Bruni, G., Rennekamp, A.J., Velenich, A., McCarroll, M., Gendelev, L., Fertsch, E., Taylor, J., Lakhani, P., Lensen, D., Evron, T., et al. (2016). Zebrafish behavioral profiling identifies multitarget antipsychotic-like compounds. Nat. Chem. Biol. *12*, 559–566.
27. Rodríguez-Pérez, R., and Bajorath, J. (2022). Evolution of support vector machine and regression modeling in chemoinformatics and drug discovery. J. Comput. Aided Mol. Des. *36*, 355–362.
28. Dara, S., Dhamercherla, S., Jadav, S.S., Babu, C.M., and Ahsan, M.J. (2022). Machine learning in drug discovery: a review. Artif. Intell. Rev. *55*, 1947–1999.
29. Chandak, T., Mayginnes, J.P., Mayes, H., and Wong, C.F. (2020). Using machine learning to improve ensemble docking for drug discovery. Proteins *88*, 1263–1270.
30. Stanley, M., Bronskill, J.F., Maziarz, K., Misztela, H., Lanini, J., Segler, M., Schneider, N., and Brockschmidt, M. (2021). Fs-mol: A Few-Shot Learning Dataset of Molecules. In Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2).
31. Vella, D., and Ebejer, J.-P. (2023). Few-shot learning for low-data drug discovery. J. Chem. Inf. Model. *63*, 27–42.
32. Chang, J., Zou, S., Xu, S., Xiao, Y., and Zhu, D. (2024). Screening of inhibitors against idiopathic pulmonary fibrosis: few-shot machine learning and molecule docking based drug repurposing. Curr. Comput. Aided Drug Des. *20*, 134–144.
33. Bontonou, M., Lioi, G., Farrugia, N., and Gripon, V. (2021). Few-shot Decoding of Brain Activation Maps. In 2021 29th European Signal Processing Conference (EUSIPCO) (IEEE), pp. 1326–1330.
34. Li, D., Ortega, P., Wei, X., and Faisal, A. (2021). Model-agnostic Meta-Learning for EEG Motor Imagery Decoding in Brain-Computer-Interfacing. In 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER) (IEEE), pp. 527–530.
35. Parng, C., Seng, W.L., Semino, C., and McGrath, P. (2002). Zebrafish: a preclinical model for drug screening. Assay Drug Dev. Technol. *1*, 41–48.
36. Lin, X., Wang, S., Yu, X., Liu, Z., Wang, F., Li, W.T., Cheng, S.H., Dai, Q., and Shi, P. (2015). High-throughput mapping of brain-wide activity in awake and drug-responsive vertebrates. Lab Chip *15*, 680–689.
37. Franceschi, L., Frasconi, P., Salzo, S., Grazzi, R., and Pontil, M. (2018). Bilevel Programming for Hyperparameter Optimization and Meta-Learning. In International conference on machine learning, pp. 1568–1577.
38. Yacouby, R., and Axman, D. (2020). Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models. In Proceedings of the first workshop on evaluation and comparison of NLP systems, pp. 79–91.
39. Carrington, A.M., Manuel, D.G., Fieguth, P.W., Ramsay, T., Osmani, V., Wernly, B., Bennett, C., Hawken, S., Magwood, O., Sheikh, Y., et al. (2023). Deep ROC Analysis and AUC as Balanced Average Accuracy, for Improved Classifier Selection, Audit and Explanation. IEEE Trans. Pattern Anal. Mach. Intell. *45*, 329–341.
40. Foody, G.M. (2020). Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification. Remote Sensing of Environment *239*, 111630.
41. Breiman, L. (2001). Machine Learning, Volume 45, Number 1 - SpringerLink. Mach. Learn. *45*, 5–32.
42. Cutler, D.R., Edwards, T.C., Jr., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., and Lawler, J.J. (2007). Random Forests for Classification in Ecology. Ecology *88*, 2783–2792.
43. Zhu, N., Zhu, C., Zhou, L., Zhu, Y., and Zhang, X. (2022). Optimization of the Random Forest Hyperparameters for Power Industrial Control Systems Intrusion Detection Using an Improved Grid Search Algorithm. Appl. Sci. *12*, 10456.
44. Sumathi, B. (2020). Grid Search Tuning of Hyperparameters in Random Forest Classifier for Customer Feedback Sentiment Prediction. Intern. J. Adv. Comput. Sci. Appl. *11*, 1.
45. Kim, T.K. (2015). T test as a parametric statistic. Korean J. Anesthesiol. *68*, 540–546.
46. Taheri, S.M., and Hesamian, G. (2013). A generalization of the Wilcoxon signed-rank test and its applications. Stat. Papers *54*, 457–470.
47. Diener, M.J. (2010). Cohen's D (The Corsini encyclopedia of psychology), p. 1.
48. McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint at arXiv. https://doi.org/10.48550/arXiv.1802.03426.
49. Armstrong, M.J., and Okun, M.S. (2020). Diagnosis and Treatment of Parkinson Disease: A Review. JAMA *323*, 548–560.
50. Jankovic, J. (2008). Parkinson's disease: clinical features and diagnosis. J. Neurol. Neurosurg. Psychiatry *79*, 368–376.
51. Jamal, M.A., and Qi, G.-J. (2019). Task Agnostic Meta-Learning for Few-Shot Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11719–11727.
52. Finn, C., Abbeel, P., and Levine, S. (2017). Model-agnostic Meta-Learning for Fast Adaptation of Deep Networks. In International Conference on Machine Learning, pp. 1126–1135.
53. Elsken, T., Staffler, B., Metzen, J.H., and Hutter, F. (2020). Meta-learning of Neural Architectures for Few-Shot Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12365–12375.
54. Liu, R., Mu, P., Yuan, X., Zeng, S., and Zhang, J. (2023). A general descent aggregation framework for gradient-based bi-level optimization. IEEE Trans. Pattern Anal. Mach. Intell. *45*, 38–57.
55. Liu, R., Mu, P., Yuan, X., Zeng, S., and Zhang, J.A. (2020). Generic First-Order Algorithmic Framework for Bi-level Programming beyond Lower-Level Singleton. In International Conference on Machine Learning (ICML) 2020.
56. Liu, R., Gao, J., Zhang, J., Meng, D., and Lin, Z. (2022). Investigating bi-level optimization for learning and vision from a unified perspective: A survey and beyond. IEEE Trans. Pattern Anal. Mach. Intell. *44*, 10045–10067.
57. Neumüller, C., Wagner, S., Kronberger, G., and Affenzeller, M. (2012). Parameter meta-optimization of metaheuristic optimization algorithms. In Computer Aided Systems Theory–EUROCAST 2011: 13th International Conference, Las Palmas de Gran Canaria, Spain, February 6-11, 2011, Revised Selected Papers, Part I 13, pp. 367–374.

58. Deng, J., Yang, Z., Ojima, I., Samaras, D., and Wang, F. (2022). Artificial intelligence in drug discovery: applications and techniques. Brief. Bioinform. *23*, bbab430.
59. Pozzan, A. (2006). Molecular descriptors and methods for ligand based virtual high throughput screening in drug discovery. Curr. Pharm. Des. *12*, 2099–2110.
60. 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. (2020). AMPL: A Data-Driven Modeling Pipeline for Drug Discovery. J. Chem. Inf. Model. *60*, 1955–1968.
61. Low, Z.Y., Farouk, I.A., and Lal, S.K. (2020). Drug Repositioning: New Approaches and Future Prospects for Life-Debilitating Diseases and the COVID-19 Pandemic Outbreak. Viruses *12*, 1058.
62. Jourdan, J.P., Bureau, R., Rochais, C., and Dallemagne, P. (2020). Drug repositioning: a brief overview. J. Pharm. Pharmacol. *72*, 1145–1151.
63. Liu, C.T., Kao, L.T., Shih, J.H., Chien, W.C., Chiu, C.H., Ma, K.H., Huang, Y.S., Cheng, C.Y., Shiue, C.Y., and Li, I.H. (2019). The effect of dextromethorphan use in Parkinson's disease: A 6-hydroxydopamine rat model and population-based study. Eur. J. Pharmacol. *862*, 172639.
64. Sebastiani, G., Almeida-Toledano, L., Serra-Delgado, M., Navarro-Tapia, E., Sailer, S., Valverde, O., Garcia-Algar, O., and Andreu-Fernández, V. (2021). Therapeutic Effects of Catechins in Less Common Neurological and Neurodegenerative Disorders. Nutrients *13*, 2232.
65. McGonigle, P. (2014). Animal models of CNS disorders. Biochem. Pharmacol. *87*, 140–149.
66. Chiang, M., Back, H.M., Lee, J.B., Oh, S., Guo, T., Girgis, S., Park, C., Haroutounian, S., and Kagan, L. (2020). Pharmacokinetic Modeling of the Impact of P-glycoprotein on Ondansetron Disposition in the Central Nervous System. Pharm. Res. (N. Y.) *37*, 205.
67. Gobena, S., Admassu, B., Kinde, M.Z., and Gessese, A.T. (2024). Proteomics and Its Current Application in Biomedical Area: Concise Review. Sci. World J. *2024*, 4454744.
68. Halder, A., and Drummond, E. (2024). Strategies for translating proteomics discoveries into drug discovery for dementia. Neural Regen. Res. *19*, 132–139.
69. Delpierre, C., and Lefèvre, T. (2023). Precision and personalized medicine: What their current definition says and silences about the model of health they promote. Implication for the development of personalized health. Front. Sociol. *8*, 1.
70. Ahrens, M.B., Orger, M.B., Robson, D.N., Li, J.M., and Keller, P.J. (2013). Whole-brain functional imaging at cellular resolution using light-sheet microscopy. Nat. Methods *10*, 413–420.
71. Westerfield, M. (1995). The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish (Brachydanio Rerio) (University of Oregon press).
72. Mauri, A. (2020). alvaDesc: A Tool to Calculate and Analyze Molecular Descriptors and Fingerprints (Springer).
73. Kingma, D.P., and Ba, J. (2014). Adam: A Method for Stochastic Optimization. Preprint at arXiv. https://doi.org/10.48550/arXiv.1412.6980.
74. Franceschi, L., Donini, M., Frasconi, P., and Pontil, M. (2017). Forward and Reverse Gradient-Based Hyperparameter Optimization. In International Conference on Machine Learning (PMLR), pp. 1165–1173.
75. Zimmerman, D.W. (1997). Teacher's Corner: A Note on Interpretation of the Paired-Samples t Test. J. Educ. Behav. Stat. *22*, 349–360.
76. Wilcoxon, F. (1992). Individual Comparisons by Ranking Methods. In Breakthroughs in Statistics: Methodology and Distribution, S. Kotz and N.L. Johnson, eds. (Springer New York), pp. 196–202.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
trihexyphenidyl	Sigma	ATC code: N04AA01
biperiden	Sigma	N04AA02
metixene	Sigma	N04AA03
procyclidine	Sigma	N04AA04
profenamine	Sigma	N04AA05
benzatropine	Sigma	N04AC01
levodopa	Sigma	N04BA01
amantadine	Sigma	N04BB01
bromocriptine	Sigma	N04BC01
pergolide	Innochem	N04BC02
ropinirole	Sigma	N04BC04
pramipexole	Sigma	N04BC05
cabergoline	Sigma	N04BC06
apomorphine	Sigma	N04BC07
piribedil	Sigma	N04BC08
rotigotine	Sigma	N04BC09
selegiline	Sigma	N04BD01
rasagiline	Sigma	N04BD02
safinamide	Sigma	N04BD03
tolcapone	Sigma	N04BX01
entacapone	Sigma	N04BX02
budipine	Sigma	N04BX03
Deposited data		
Analyzed BAM data	10.5281/zenodo.13362829	n01-n07 BAM
AlvaDesc calculated CDA data	10.5281/zenodo.13362829	n01-n07 CDA
Biological target database	Drugbank	https://go.drugbank.com/targets
Main code for *Meta*-CNN algorithm	10.5281/zenodo.13362829	meta-learning-BAM.py; README.md
Experimental models: organisms/strains		
Zebrafish Larvae	Champalimaud Research	elval3:GCaMP5G
Software and algorithms		
ImageJ	LOCI, University of Wisconsin	RRID:SCR_003070
AlvaDesc	Alvascience	https://www.alvascience.com/alvadesc/
Gradient-based hyperparameter optimization and meta-learning package	FAR-HO	https://github.com/lucfra/FAR-HO

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Zebrafish line

Adult transgenic zebrafish line elval3:GCaMP5G (Champalimaud Research)[70] were cultivated according to standard laboratory protocol,[71] embryos were harvested and nurtured in E3 medium (5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$, and 0.33 mM $MgSO_4$) at a stable temperature of 28°C. Pigmentation was inhibited by treatment with 0.003% 1-phenyl 2-thiourea (PTU, Sigma). All animal experimental procedures were conducted with prior approval from the City University of Hong Kong's animal ethics committee and complied with local animal care guidelines.

METHOD DETAILS

Microfluidic immobilization and drug exposure

Zebrafish larvae aged 6–8 days post-fertilization (dpf) were employed in high-throughput brain activity mapping experiments. A microfluidic chip was fabricated using molded polydimethylsiloxane (PDMS) and glass substrates.[24] Larvae were transferred and immobilized utilizing an automated system, which incorporated fluidic circuitry, electromagnetic valves, and a video detection apparatus.[24] The BAM and CDA library included 331 drugs, classified according to the WHO ATC system. During the experiment, each drug was dissolved in dimethyl sulfoxide (DMSO, Sigma) at a standard concentration of ~10 mM, which was then diluted to a working concentration of ~10 μM for experimental application. For each chemical treatment, we employed five biological replicates. Each replicate was subjected to identical experimental conditions to ensure consistency and reproducibility. Control groups, where no treatment was applied, were included in each experiment to establish baseline brain activity and distinguish treatment effects from background noise. To minimize pigment variations, all fishes were uniformly treated with 1-phenyl 2-thiourea (PTU). Environmental factors such as temperature, light, and feeding schedules were controlled and kept consistent across all experiments.

Image acquisition and T-score BAM generation

Larvae immobilized on the microfluidic chip underwent automated imaging via an inverted fluorescence microscope (Olympus IX81) equipped with a 10x objective (NA, 0.4). A sequence of 4000 frames per larva (dual-layer focus, 150 ms exposure time, no shutter pause) was recorded. Image stacks were pre-processed with an ImageJ macro script to mitigate motion artifacts, any blurred frames were excised. The TurboReg plugin and defined regions of interest (ROIs) facilitated the creation of concentric XY-cropped montage images, allowing for the assessment of calcium transients across specimens. Image preprocessing methods were derived from our previous study.[24,36] The alignment of brain images to a standard template was conducted using a custom-made computational pipeline. This pipeline consists of the following steps: 1. Consistent Imaging Settings: during the acquisition stage, all imaging settings were maintained consistently to minimize any systematic variations. 2. Fluorescence Region Extraction: the dark background was removed from each raw image to isolate the fluorescence brain region. 3. Template Mapping: the fluorescence brain region was mapped to the standard template through specific transformative adjustments (e.g., rotation, translation) using the brain center-line as a registration landmark. This step maximized symmetry with respect to the center line in the resultant image. 4. Region Isolation: the fish eye regions were further removed by retaining only the regions within the template for downstream analysis. After the alignment, brain images from each fish were resized and rotated to ensure uniform orientation and scale (Figure S2). Each frame collected from calcium imaging was divided into small regions of interest (ROIs), each covering an area of 15.21 μm^2. The fluorescence signal time series in each ROI was filtered using a high-pass filter with a cut-off frequency of 0.2 Hz to remove illumination fluctuation interference. The filtered fluorescence signal was used to calculate calcium transients, which indicate neural activity. A threshold, defined as the mean plus two times the standard deviation of the ΔF/F curve (change in fluorescence over baseline fluorescence), was applied to detect significant changes in calcium levels. Calcium transients were counted over a 10-min period before and a 15-min period after compound treatment. A BAM matrix $A = [aij]$ was generated by calculating the change in calcium transient counts in each ROI before ($t0$) and after ($t1$) chemical treatment, summed across multiple layers along the z axis.[24,36]

$$a_{ij} = \sum_{k} \left(c_{ijk}^{t1} - c_{ijk}^{t0} \right)$$

where c_{ijk}^{t0} and c_{ijk}^{t1} represent the calcium transient counts in the ROI at row i and column j of the k-th layer before and after treatment, respectively. T-scores were calculated to quantify the brain activity differences between conditions and a standard template fitting was performed to align the images to a common reference.

Molecular descriptor database

Three-dimensional molecular descriptors were calculated utilizing alvaDesc,[72] a software that computes model-based physicochemical properties of molecules, such as topological polar surface area (TPSA), molar refractivity, and pharmacophore descriptors, which are related to the chemical properties of compounds. To extract comprehensive information from the molecular structure, 3D structures were employed. The original 3D spatial data file (SDF) conformer files for the experimental chemicals were sourced from PubChem. The OCHEM online educational-free database facilitated the calculation of alvaDesc v1.0.16 molecular descriptors, which encompasses 5,305 descriptors derived from 3D data. Furthermore, the molecular structures were refined using the OpenBabel optimization tool throughout the computational process.

The data pre-processing workflow

Image pre-processing steps contained resizing, rotation, brain activity maps generation, T-score maps calculation, and uniform template fitting (Figure S1A). Subsequently, a principal component analysis (PCA) algorithm was applied to reduce dimensionality and identify patterns. Each T-score BAM was then converted into a dimensionality-reduced Pheno-Print, which was described by a 20-dimensional vector (Figure S2). For the BAM library, an unsupervised classification algorithm was utilized to define functional clusters. Optimal phenotypic clustering was determined using CDF curves from consensus clustering results, revealing that ten phenotypic BAM clusters provided the best

feature extraction. The process paralleled the CDA analysis, where descriptor arrays replaced T-score BAMs. AlvaDesc was employed to compute the 3D structures of each drug as molecular descriptors and footprints, resulting in 5305 features.[72]

Features that were PCA outliers or had a median absolute deviation (MAD) ≤ 0.5 after standard normalization were removed to reduce dimensionality and focus on relevant features. For both BAM and CDA datasets, principal component analysis (PCA) was employed to reduce the dimensionality of the data while retaining the most significant variance. For BAM, the top 20 principal components explained 53% of the total variance and were used for further analysis (Figure S1). For CDA, PCA was applied post-feature selection to ensure that only the most informative features were used in clustering. Based on the consensus clustering of the PCA-reduced data, ten phenotypic BAM clusters were identified, guided by the CDF curve. Similarly, the molecular descriptors of the drugs were clustered into ten groups using consensus clustering, ensuring robust grouping based on molecular features. Stringent quality control measures were applied, including the removal of PCA outliers and features with low variability (MAD ≤ 0.5), to ensure high-quality and reliable data for clustering and analysis. The correlation between BAM/CDA clusters and WHO functional drug categories was quantified using the hypergeometric test. This statistical method assessed the robustness of the associations, which were visually depicted in a heatmap (Figure S1C) This graphical illustration confirmed the correspondence between clustered features and drug categories.

Data partition

In the CNS ATC-BAM correlation experiments, T-score BAMs were annotated with corresponding secondary ATC codes, such as N0X, denoting distinct pharmacological effects (N01: Anesthetics, N02: Analgesics, N03: Antiepileptics, N04: Anti-Parkinson Drugs, N05: Psycholeptics, N06: Psychoanaleptics, N07: Other Nervous System Drugs). To assess the *Meta*-CNN model's capability in CNS drug categorization, we employed the control experiment of a CNN-only model with the same framework. The datasets were categorized into seven subgroups reflecting the CNS drug categories as defined by WHO. Distinct data partitioning strategies were adopted for the *Meta*-CNN and the conventional CNN models: the former's dataset was segmented into meta-training and meta-testing sets, encompassing training/validation and training/testing subsets, respectively. In contrast, the latter utilized only training and testing sets. Based on the outcomes from various data splits, the optimal division for the CNN was determined as 18 training samples and 3 testing samples. For the *Meta*-CNN, 4 samples were allocated for meta-training, 3 for validation, 11 for meta-testing, and 3 for final testing (Table S3). The quantity of samples was dictated by the minimum number observed across the groups (Table S1). To ensure comprehensive evaluation given our dataset size and to maintain learning stability, a 4-fold cross-validation was implemented during the training phase (M = 4), and an 8-fold cross-validation during the testing phase (Q = 8), which ensured the near-complete coverage of the dataset, approximating 100% traversal. An epoch count of 200 was selected to strike a balance between learning efficiency and the risk of overfitting, with a nearly zero loss in training progress (Figure 3A).

Meta-learning algorithm design

The meta-learning process is bifurcated into two phases: meta-training and meta-testing (Figures 5B and 5C). Following the experimental protocol used in the recent works,[54,55] we separated the network architecture into two parts: the cross-task intermediate representation layers (parameterized by ω), which output the meta-features, and the multinomial logistic regression layer (parameterized by θ^i), which served as our ground classifier for the i-th task. During the meta-training phase, we divided the training sample set into training and validation sets. This allows us to extract common characteristics of drugs, known as the hyper-parameter ω^*. We tested M randomly repeated meta-training tasks to learn ω^* by selecting data from the meta-training set. Thus, the meta-training collection is represented as $D_{m-train} = \{(D^{train}_{m-train}, D^{val}_{m-train})^{(i)}\}^{M}_{i=1}$, where $(D^{train}_{m-train}, D^{val}_{m-train})^{(i)}$ represents the training and validation data for the i-th task in the meta-training phase. Based on the segmented dataset, we establish a bilevel optimization model to learn the hyper-parameter ω^* across the M tasks. Then for the i-th task, we consider the cross-entropy function as the task-specific loss, and thus the lower-level objective can be defined as:

$$f(\omega, \theta^i) := \sum_{i=1}^{M} L\left(\omega, \theta^i; D^{train(i)}_{m-train}\right) \quad \text{(Equation 1)}$$

where θ^i is the general parameter vector of the i-th task. For the upper-level objective, we also use the cross-entropy function, but define it based on $D^{val}_{m-train}$ as

$$F(\omega, \theta) := \sum_{i=1}^{M} L\left(\omega, \theta^{*(i)}; D^{val(i)}_{m-train}\right) \quad \text{(Equation 2)}$$

where ω is the hyper-parameter vector. The objective is to learn the hyper-parameter ω^* across tasks in the meta-training loop and solve the ensuing bilevel optimization challenge encompassing all tasks (Figure 5A):

$$min_{\omega,\theta}\left\{F(\omega, \theta) = \sum_{i=1}^{M} L\left(\omega, \theta^{*(i)}; D^{val(i)}_{m-train}\right)\right\} \quad \text{(Equation 3)}$$

$$\text{s.t. } \theta^{*(i)} \in argmin_{\theta^i}\left\{f(\omega, \theta^i) = \sum_{i=1}^{M} L\left(\omega, \theta^i; D^{train(i)}_{m-train}\right)\right\} \quad \text{(Equation 4)}$$

where L denotes the cross-entropy loss function.

For the lower-level problem, starting from the initialization point $\theta_0^{(i)}$, we generated a sequence $\{\theta_{k+1}^{(i)}\}_{k=0}^{K}$ parameterized by ω as $\theta_{k+1}^{(i)}(\omega) = \mathcal{T}_{k+1}(\omega, \theta_k^{(i)}(\omega))$, where $k = 1, 2, ..., K$ represents the iterations. The formulate $\mathcal{T}_{k+1}$ aggregates the descent information of F and f. Specifically, for a given ω, the descent directions of the upper-level and lower-level objectives can be defined as

$$\mathcal{T}_{k+1}\left(\omega, \theta_k^{(i)}(\omega)\right) = \theta_k^{(i)} - \left(\alpha_k \nabla_{\theta^i} F(\omega, \theta) + \beta_k \nabla_{\theta^i} f\left(\omega, \theta^i\right)\right) \quad \text{(Equation 5)}$$

where α_k and β_k denote the aggregation parameters. After obtaining the $\theta_K^{(i)}(\omega)$ after K iterations, we incorporated $\theta_K^{(i)}(\omega)$ into the upper-level objective and obtained a single-level approximation model $min_\omega\{F(\omega, \theta_K^{(i)}(\omega))\}$. The Adaptive Moment Estimation (Adam) optimization algorithm, based on automatic differentiation techniques, is then used to achieve optimal solutions to the single-level approximation problem. This algorithm exhibits relatively stable optimization convergence properties.[52]

The neural network models

Both the *Meta*-CNN and CNN-only models are structured around pairs of multi-layer CNNs, consisting predominantly of convolutional layers followed by a fully connected layer. The distinction between the two models lies in the fact that the CNN-only model does not incorporate the meta-learning framework for deriving hyperparameters; it solely relies on the CNN framework. This architecture is intricately designed to extract and process features from the input data. Optimizing the parameters of each CNN model is a critical aspect, involving diverse mathematical models and algorithms for training and fine-tuning. These algorithms are essential as they iteratively update the weights and biases within the models, enhancing their ability to fit the training data and minimize prediction errors accurately. Therefore, the choice of these algorithms is a key factor in determining the models' classification performance. In our discussion on model algorithms and network construction, we delved into the two phases of meta-learning. The meta-training phase utilizes the CNN model to derive hyper-parameters, following the methodology,[37] the network architecture here is bifurcated into two segments: the inter-task intermediate representation layers, functioning as the hyper-parameter ω for meta-feature generation, and the multinomial logistic regression layer, which articulates the representation mapping as the general network parameter θ^i for classifying the i-th task. The optimization of hyper-parameters in this setup necessitates efficient algorithms capable of handling a dual-layer optimization model. For this purpose, our study employed the Adaptive Moment (Adam) optimization algorithm for the upper-level objective $F(\omega, \theta)$.[73] While for the lower-level subproblem $f(\omega, \theta^i)$, we applied previous study about Bidirectional Approximate Gradient (BDA) algorithm,[55,56] which is noted for its superior convergence performance and generalization capability compared with other algorithms.[37] In the meta-testing phase, the secondary CNN model is employed to obtain the general network parameters and predict classification outcomes. Leveraging the learned hyper-parameter ω^* from the first phase, we applied the stochastic gradient algorithm to train the general network parameter $\theta^{*(j)}$ for each j-th task. The foundational settings of both CNN models are consistent, with the representation mapping implemented via a 4-layer CNN equipped with stridden convolutions and 64 filters per layer. These two phases are integrated into a unified CNN model for predicting classifications on the test set. All parameters are meticulously optimized by minimizing the cross-entropy loss function. Our meta-learning framework, through the synergistic combination of a bi-level programming algorithm and a CNN model, achieves rapid adaptation and precise classification in scenarios characterized by few-shot learning.

Code package

Our methodology builds upon the FAR-HO package, a gradient-based hyperparameter optimization and meta-learning framework.[74] This package provides a unified mathematical framework grounded in bilevel programming, which effectively integrates gradient-based hyperparameter optimization and meta-learning techniques. All code was implemented in Python, utilizing TensorFlow and various associated libraries, the far_ho library was employed for hyperparameter optimization. The framework was designed to handle various tasks, adapting to new tasks with limited data via meta-training and meta-testing cycles.

The model architecture was constructed using a CNN, which served as the feature extractor across tasks. The CNN was composed of four convolutional layers, designed to capture task-specific representations efficiently. The feature representations were flattened using tf.reshape function to form a suitable input for the subsequent fully connected layers. The classifier was implemented using TensorFlow Contrib's tcl.fully_connected function, where the final layer output was dimensioned to match the number of classes per task, and weights were initialized to zeros. This structure allowed the network to learn task-specific classifiers during the meta-training phase. The meta-training phase involved generating meta-batches from the dataset, tailored for few-shot learning tasks. Each meta-batch contained a set of tasks, where the model was trained on a small sample of data and evaluated on a separate validation set to simulate the few-shot learning scenario. Hyperparameter optimization was critical to the meta-learning process. We utilized far_ho.HyperOptimizer to perform a bi-level optimization, where the inner loop consisted of a task-specific gradient descent optimizer, and the outer loop was optimized using the Adam optimizer. The hyperparameters, including learning rates and layer weights, were optimized over 200 meta-iterations, enhancing the model's ability to generalize across tasks. Meta-evaluation was conducted by testing the model on unseen tasks, measuring performance using metrics such as accuracy and precision. These metrics were computed using the scikit-learn library. The evaluation focused on how well the model adapted to new tasks, reflecting its meta-learning capabilities. The results of the meta-evaluation were aggregated across tasks to compute the average performance, which was then used to assess the efficacy of the meta-learning process.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses and hypothesis testing procedures

To evaluate the performance differences between various models, we applied a detailed set of statistical analyses and hypothesis testing procedures, including paired t-tests, Wilcoxon signed-rank tests, and Cohen's d effect size calculations. Paired t-tests were used to determine if there was a significant difference between the means of two related groups. The assumptions made for this test included that the differences between paired observations were normally distributed and that the pairs were randomly and independently selected from the population.[75] The procedure involved calculating the mean difference $\overline{d}$ and the standard deviation (SD) of the differences, computing the t-statistic using $t = \frac{\overline{d}}{SD/\sqrt{n}}$ where n is the number of pairs, and determining the p-values by comparing these statistics against the critical values from the t-distribution with $n-1$ degrees of freedom. To complement the paired t-tests, Wilcoxon signed-rank tests were performed as a non-parametric alternative that does not assume a normal distribution of the differences.[76] This test assumed that the pairs were randomly and independently selected and that the differences between pairs were symmetrically distributed about the median. The procedure involved calculating the differences between paired samples, ranking the absolute values of these differences, summing the ranks for positive and negative differences, and using the smaller rank sum (min (T^+, T^-)) as the test statistic. The p-value was then determined by comparing this test statistic to critical values from the Wilcoxon distribution. For quantifying the magnitude of differences, Cohen's d was calculated.[47] This involved gathering samples from the two groups to be compared, computing the means ($\overline{X_1}$ and $\overline{X_2}$) and SD for each group, calculating the pooled standard deviation $SD_{pooled} = \sqrt{\frac{(SD_1^2, SD_2^2)}{2}}$, and using the formula $d = \frac{\overline{X1} - \overline{X_2}}{SD_{pooled}}$. Cohen's d values were interpreted as small (0.2), medium (0.5), or large (0.8) effect sizes.