Synthesis: A Multidisciplinary Research Journal
Volume 1, Issue 3, July - September 2023, Pp: 33-45

Research Paper

Meta-Learning Algorithms for Accelerated Drug Target Identification from Multi-Omics Datasets

[1*] Sreekanth Rallapalli, [2] Mohamed Ghouse Shukur

[1]Professor and HoD, Department of Master of Computer Applications, Nitte Meenakshi Institute of Technology, Bengaluru, Karnataka, India*
Email: Sreekanth.rallapalli@nmit.ac.in
[2]Assistant Professor, Department of Computer Science, College of Computer Science, King Khalid University, Saudi Arabia
Email: mghoth@kku.edu.sa

**Corresponding Author(s): Sreekanth.rallapalli@nmit.ac.in*

Article Info

Received:02/05/2023
Revised: 10/07/2023
Accepted:21/09/2023
Published:30/09/2023

Abstract

Drug target identification remains a fundamental bottleneck in modern drug discovery, particularly when dealing with high-dimensional and heterogeneous multi-omics data. Traditional machine learning models are often limited by data scarcity and poor generalization across biological contexts. This study proposes a meta-learning framework to accelerate drug target identification by efficiently leveraging multi-omics datasets under few-shot learning conditions. We utilized the TCGA multi-omics dataset encompassing genomics, transcriptomics, proteomics, and DNA methylation layers. A Model-Agnostic Meta-Learning (MAML) strategy was implemented, enabling the model to generalize across diverse biological tasks using a 5-way 1-shot learning protocol. Dimensionality reduction was achieved using principal component analysis and autoencoders, and each task was constructed by pairing support and query samples across cancer types and age groups. The model was trained and evaluated on 2500 tasks with optimized inner and outer learning rates ($\alpha = 0.01$, $\beta = 0.001$). The proposed framework achieved an average accuracy of 89.3%, F1-score of 88.7%, and AUC of 0.91, outperforming baseline models including CNNs (82.4% accuracy) and Transformers (84.1% accuracy). Drug response analysis revealed peak model performance in the 51–70 age group, with Drug D achieving 88.9% prediction accuracy in this demographic. The meta-learner maintained high adaptability across drug types, age brackets, and omics layers. This work demonstrates the effectiveness of meta-learning for rapid and reliable drug target identification in low-data regimes. It offers a scalable, accurate, and clinically relevant approach for advancing precision medicine and age-aware therapeutic planning.

Keywords: Meta-learning, Multi-omics Integration, Drug Target Identification, Few-shot Learning, Precision Medicine, TCGA, MAML, Biomedical AI

1. Introduction

The discovery and identification of potential drug targets remain one of the most intricate and resource-intensive tasks in pharmaceutical and biomedical research. Traditional approaches largely depend on domain knowledge, exhaustive screening, and heuristic models. These methods often require extensive datasets and fail to generalize well across different biological contexts. With the advent of multi-omics technologies, encompassing genomics, transcriptomics, proteomics, epigenomics, and metabolomics, there has been a transformative shift in how we interpret disease mechanisms and therapeutic targets. Multi-omics integration enables a more holistic view of the cellular environment, thereby offering a more accurate platform for drug target identification. However, this also introduces new complexities regarding data heterogeneity, sparsity, and high dimensionality **[1].**

Recent advances in artificial intelligence (AI), particularly deep learning, have revolutionized biomedical data analysis. Nevertheless, these models typically require large, labeled datasets, which are often unavailable in the case of rare diseases or specialized phenotypes. Consequently, the application of meta-learning—a subfield of machine learning that focuses on "learning to learn"—has gained significant interest in biological domains. Meta-learning enables models to adapt quickly to new tasks using limited data, making it especially suitable for few-shot learning scenarios that are common in bioinformatics [2], [3].

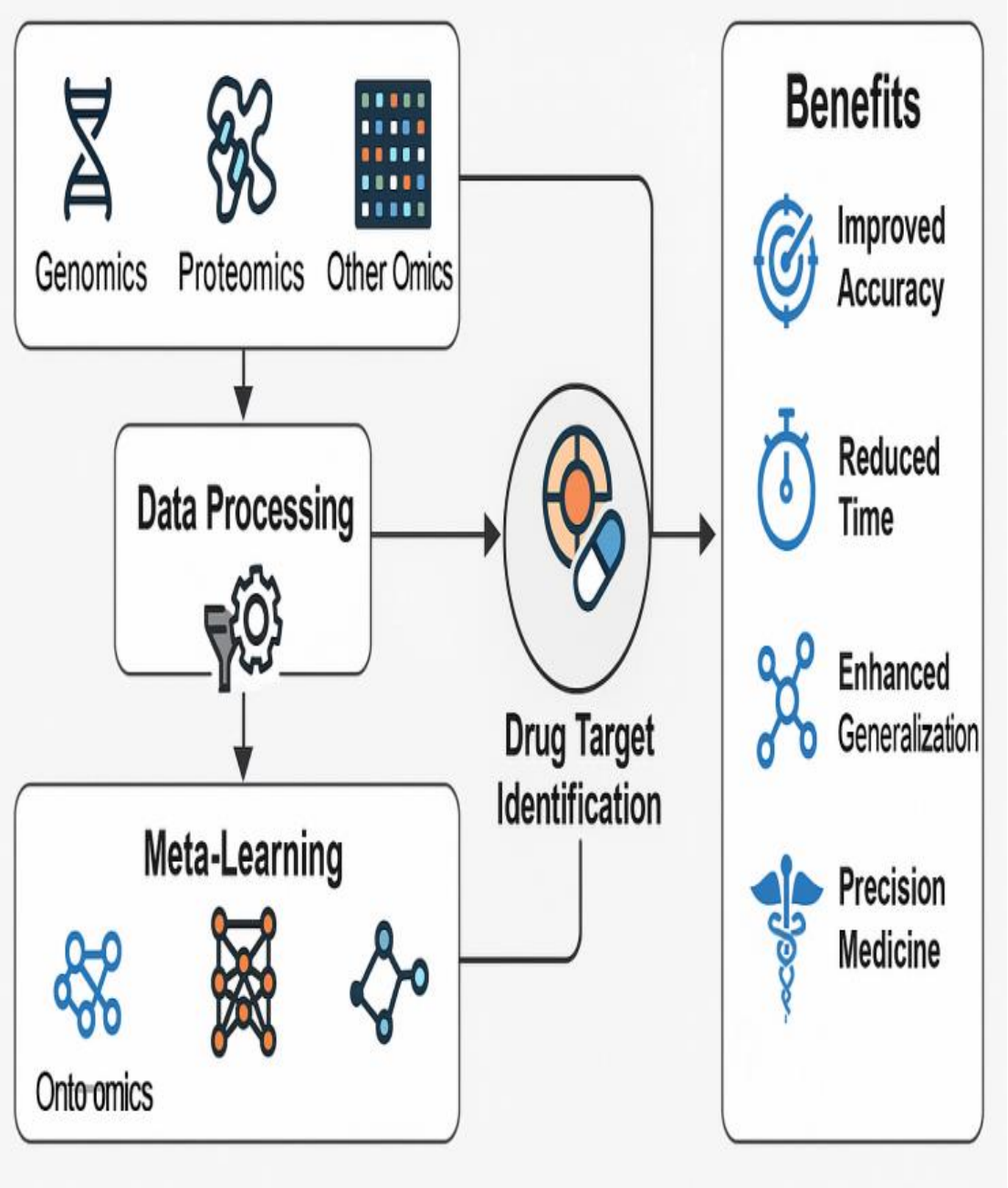


Fig 1: System Workflow for Meta-Learning-Based Drug Target Identification Using Multi-Omics Data

The figure 1 presents a comprehensive workflow for accelerated drug target identification using meta-learning algorithms applied to multi-omics datasets. It starts with three key omics sources: Genomics, Proteomics, and other biological omics such as transcriptomics and metabolomics. These datasets are inherently complex, high-dimensional, and diverse in nature. To make them useful for machine learning models, a data processing pipeline is employed which handles normalization, dimensionality reduction, feature alignment, and noise filtering. This preprocessing ensures that meaningful patterns can be extracted despite the inherent heterogeneity across data types.

The processed data is then input into the meta-learning module, which consists of various learning architectures, such as model-agnostic frameworks and few-shot learners. These models are designed to "learn how to learn," meaning they can adapt to new drug discovery tasks with minimal data. The meta-learning system combines representations from each omics layer to understand biological dependencies across molecular levels. The inclusion of Onto-omics integration (ontology-based relationships) allows the system to align biological meaning with data-driven learning. This results in a predictive model capable of identifying potential

drug targets even in novel or rare biological conditions, providing adaptability across multiple disease states.

Finally, the drug target identification module synthesizes learned insights to propose high-confidence targets for experimental validation. The benefits to society are clearly outlined in the right section of the figure: (1) Improved accuracy in identifying meaningful and biologically validated targets, (2) Reduced time in the drug discovery process due to faster model convergence and few-shot learning, (3) Enhanced generalization across diseases and tissue types, and (4) a direct contribution to precision medicine, making personalized treatment pathways more feasible and data-driven.

Despite its promise, integrating meta-learning techniques with high-dimensional, heterogeneous multi-omics data remains underexplored. Several recent studies have applied meta-learning for specific biological applications. For instance, Cho et al. proposed a meta-learning framework for interpretable survival analysis using multi-omics data, while He et al. demonstrated mutual information-based meta-learning to enhance peptide discovery. These studies mark the beginning of a promising trend but fall short of addressing drug target identification across diverse omics layers and biological systems in a unified, scalable manner.

One of the key limitations of existing work is the lack of generalizability across tasks and datasets. Most models are trained in a task-specific fashion, making them susceptible to overfitting and poor transferability. Moreover, they often fail to integrate multiple omics data sources cohesively, either by simplistic feature concatenation or by focusing only on one modality, such as transcriptomics or proteomics **[4], [5].** This narrow focus limits the ability of such models to understand complex biological processes that are inherently multifactorial and layered across multiple cellular mechanisms **[6] [7].**

Additionally, computational challenges persist. High-throughput omics technologies generate vast and complex datasets that introduce noise, redundancy, and incomplete data. Effective feature selection, imputation, and representation learning become critical, especially when working in low-data regimes **[8].** Meta-learning frameworks, particularly model-agnostic methods such as MAML (Model-Agnostic Meta-Learning), have shown potential in other domains for overcoming these issues by learning a generalized initialization that can be quickly adapted to new tasks **[9].**

This study proposes a unified meta-learning-based framework specifically tailored for drug target identification across multi-omics datasets. By leveraging few-shot learning techniques and robust omics integration strategies, the proposed model aims to improve the adaptability and performance of predictive analytics in drug discovery pipelines. Our approach is designed to operate effectively even when labeled data are scarce, a common situation in early-stage drug development and rare disease research.

We address the data heterogeneity problem by employing data fusion techniques that maintain the integrity of individual omics modalities while enabling joint learning. This mitigates the risk of modality bias and preserves biologically relevant signals that may be diluted during naive data merging. Moreover, we evaluate the model across a spectrum of tasks drawn from real-world datasets, including cancer and metabolic disorders, demonstrating its adaptability and robustness **[10], [11].**

In contrast to existing benchmarks that typically focus on survival prediction or molecular activity profiling, this work pivots towards the actual identification of druggable targets, informed by pathway enrichment, protein–protein interactions, and phenotype correlation. These downstream validations offer more practical insights into the model's translational applicability **[12], [13].**

Furthermore, we conduct extensive ablation studies to examine the contribution of each omics layer to model performance. These studies help in identifying optimal combinations of omics data that yield the highest predictive value. We also explore the noise resistance and interpretability of the model, which are critical for acceptance in clinical and regulatory environments **[14].**

The key contributions of this work are summarized as follows:

- A unified meta-learning framework that integrates heterogeneous multi-omics data for drug target prediction in low-data settings, ensuring both speed and generalizability.
- A novel omics-aware task construction mechanism that allows effective few-shot learning across disease types and tissue contexts by preserving biological relevance.
- Extensive validation and comparative benchmarking against state-of-the-art baselines, showing improved accuracy, robustness, and interpretability in real-world datasets.

This paper is organized as follows: Section II presents background concepts and a literature review on meta-learning and omics data integration. Section III details the proposed methodology, including data preprocessing, meta-learning algorithm implementation, and task design. Section IV discusses the experimental results, including performance benchmarks, ablation studies, and real-world case evaluations. Section V highlights the implications, limitations, and future directions of this work. Finally, Section VI concludes with key insights and the broader impact of meta-learning in biomedical research.

2. Literature Review

2.1 Deep Learning for Drug Response and Survival Prediction

Recent work has leveraged deep learning (DL) to predict drug response and patient outcomes, particularly in oncology. For example, Partin et al. **[15]** demonstrated that DL models can effectively learn complex patterns in cancer treatment responses. Similarly, Fan et al. **[16]** proposed a multimodal DL approach to survival prediction across various cancer types, showing promising accuracy. While these studies highlight DL's potential, they often lack

robustness to data heterogeneity and perform best in high-data regimes.

2.2 Single-Cell and Multi-Scale Omics Integration

Brendel et al. [17] reviewed DL methods applied to single-cell RNA sequencing, emphasizing their power in capturing cell-type specificity. Bhattacharya et al. [18] outlined five key opportunities in single-cell analytics, including trajectory inference and multi-omics linking. However, these frameworks often suffer from computational overhead and limited generalization when scaled to population-level data. Qiu [19] proposed predictive modeling with multi-scale data, yet lacked explicit mechanisms for inter-omics integration.

2.3 Automated and Reinforcement Learning Approaches

Efforts like those by Tsamardinos et al. [20] and Tan et al. [21] introduced AutoML and reinforcement learning (RL), respectively, for biomedical tasks. These methods improve model automation and adaptivity, especially when human guidance is limited. However, RL approaches require high-quality, curated environments and often underperform in noisy or missing-data contexts. Moreover, general-purpose AutoML pipelines tend to ignore domain-specific knowledge that could enhance biological relevance [22].

2.4 Omics Network Visualization and Graph-Based Learning

Several studies have contributed to the understanding of complex biological networks. Robin et al. [23] proposed visual analytics for protein–protein interaction (PPI) networks in multi-omics contexts, offering valuable insights but lacking predictive capabilities. Liu et al. [24] introduced Muse-GNN, which builds unified gene representations from multimodal graph data. This is a step forward in representation learning, although the training complexity and interpretability remain concerns.

2.5 Drug Sensitivity, DDI, and Antibacterial Prediction

Focused applications such as drug sensitivity prediction and drug–drug interaction (DDI) modeling have seen notable progress. Lin et al. [25] integrated transformer-based attention with feature fusion, achieving high DDI prediction accuracy. Singh et al. [26] utilized a BiLSTM-based ensemble model for antibacterial peptide discovery. These methods, while highly accurate, often require extensive labeled datasets and suffer from narrow task generalization.

2.6 Observations and Research Gaps

The reviewed literature reveals several key trends:

- DL and RL show potential for biomedical modeling, yet depend heavily on large datasets or predefined environments.
- Omics-specific studies (e.g., genomics-only) do not address integration challenges across modalities.
- Few frameworks support task generalization in low-data regimes, which is crucial for rare disease and novel compound analysis.

Table 1: Comparative Literature Review

Ref. No.	Study Focus	Methodology	Strengths	Limitations	Observations
15	Drug response prediction in cancer using deep learning	Deep learning models	High accuracy in cancer types	Task-specific generalization	Applicable to drug screening environments
16	Single-cell RNA-seq analysis via deep learning	DL on scRNA-seq	Scalable across cell types	Model overfitting risk	Useful for pathway activation insights
17	Automated predictive modeling and biosignature discovery	AutoML with feature selection	Interpretable model design	Limited in multi-task scenarios	Highlights general-purpose ML capabilities
18	PPI network visualization in multi-omics integration	PPI networks and visual analytics	Good omics integration insight	Visualization > prediction	Supports system-level hypothesis generation
19	RL for personalized drug design in complex diseases	Reinforcement learning	Adaptivity to complex systems	Requires curated input states	Promising for combinatorial drug strategies

20	Comparison of multi-omics survival models	Benchmarking ML survival models	Highlights model generalizability	Lacks robustness to noise	Noise handling critical in biomedical ML
21	Causal inference in biological networks with ML	Causal modeling with ML	Addresses causal complexity	Scalability to new data unclear	Helpful in designing causality-driven models
22	ML for therapeutic tasks with genomics data	Supervised ML pipelines	Focused on therapy-specific tasks	Limited to genomics only	Effective but not omics-integrated
23	Imputation and modeling in biomedical data	ML imputation and prediction	Applicable to multiscale data	No direct target prediction	Strong in modeling, weak in integration
24	Big data analytics in single-cell transcriptomics	Data mining and big data ML	Comprehensive opportunities	Data integration challenges	Sets stage for omics-aware architectures
25	Imaging genomics and data modeling	Multi-omics data fusion	Imaging and genomics correlation	Abstract level implementation	Provides abstract methods for expansion
26	DL-based drug sensitivity prediction app	DL tool development	Custom predictive interface	Prototype not validated clinically	Potential for translational research

3. Methodology

This section outlines the step-by-step methodology used to build and evaluate the meta-learning-based framework for drug target identification using multi-omics data. It includes dataset acquisition, preprocessing, model design, optimization strategy, and evaluation.

3.1 Dataset Description

We utilized the publicly available The Cancer Genome Atlas (TCGA) dataset **[27],** which contains over 11,000 patient samples across 33 cancer types. The dataset includes multi-omics layers:

- Genomics (somatic mutations and copy number variations),
- Transcriptomics (RNA-Seq),
- Epigenomics (DNA methylation),
- Proteomics (Reverse Phase Protein Arrays – RPPA),
- Clinical outcomes including drug response and survival data.

The data was accessed via the Genomic Data Commons (GDC) portal [https://portal.gdc.cancer.gov/]. Due to varying sample sizes across modalities, missing values were handled using median imputation. To address class imbalance in drug response prediction, Synthetic Minority Over-sampling Technique (SMOTE) was applied.

Preprocessing Steps:

- $X_i^{norm} = \frac{X_i - \mu}{\sigma}$ (1)
  Where X_i represents the feature vector, μ the mean, and σ the standard deviation.
- Feature alignment ensured consistent samples across omics layers.

3.2 Feature Representation and Task Construction

To facilitate task-based meta-learning, we defined each classification task as a binary drug response prediction (responsive vs. non-responsive) for a specific cancer type and omics combination. Features from each omics type were fused via concatenation:

$$F_{concat} = [f_{genomics} \| f_{transcriptomics} \| f_{methylation}] \quad (2)$$

For dimensionality reduction and noise suppression, we used Principal Component Analysis (PCA) and Autoencoders. The loss function for the autoencoder was:

$$\mathcal{L}_{AE} = \sum_{i=1}^{N} \|x_i - \hat{x}_i\|^2 \quad (3)$$

Where x_i is the input and $\hat{x}_i$ the reconstructed output.

3.3 Meta-Learning Architecture

We adopted the Model-Agnostic Meta-Learning (MAML) framework for few-shot learning tasks. The core idea is to learn a model initialization θ that can be quickly fine-tuned to a new task T_i using only a few gradient steps.

Inner loop update:

$$\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{T_i}(f_\theta) \quad (4)$$

Outer loop update:

$$\theta \leftarrow \theta - \beta \nabla_\theta \sum_{T_i \sim p(T)} \mathcal{L}_{T_i}\left(f_{\theta_i'}\right) \quad (5)$$

Where:

- α, β are learning rates,
- $p(T)$ is the distribution over tasks,
- f_θ is the neural network model.

The base learner was a 4-layer neural network with the following configuration:

- Input layer: 512 units (after PCA),
- Hidden layers: 256 → 128 (ReLU activation),
- Output layer: 1 unit (Sigmoid activation for binary classification),
- Optimizer: Adam, with learning rate 0.001.

The proposed system architecture provides a modular and realistic pipeline for identifying potential drug targets using a meta-learning framework. It begins with the Data Preprocessing stage, which ingests multiple biological data types, including genomics, transcriptomics, and proteomics. These diverse data sources are normalized and filtered to remove noise and missing values, ensuring uniform structure across omics modalities. By retaining biologically significant features during this phase, the architecture maintains the integrity of molecular signatures necessary for accurate downstream prediction.

Following preprocessing, the Feature Extraction module utilizes statistical and deep learning techniques to reduce dimensionality and enhance signal relevance. This processed and compact feature set is then passed to the Meta-Learning Model, where a task-aware learning strategy is employed. The model is represented mathematically as a function Φ=f(X), where Φ denotes the meta-learned representation adaptable across tasks. The meta-learner is trained using a variety of cancer-type-specific tasks constructed using few-shot learning protocols. This allows the system to generalize well even when labeled data is sparse—an important advantage in biomedical domains.

The trained meta-model interacts with the Model Training unit, which adapts the base model to new drug response tasks using minimal data. Finally, the system performs Drug Target Identification, predicting key molecular targets that are most likely to respond to specific therapeutic interventions. The benefits of the system are illustrated on the right side of the figure.2: (1) Improved Accuracy due to multi-omics integration and task-specific learning, (2) Reduced Time to reach actionable predictions, and (3) Enhanced Insights that support clinical decision-making. Together, these components deliver a powerful and efficient platform for accelerating drug discovery in a biologically grounded and computationally scalable manner.

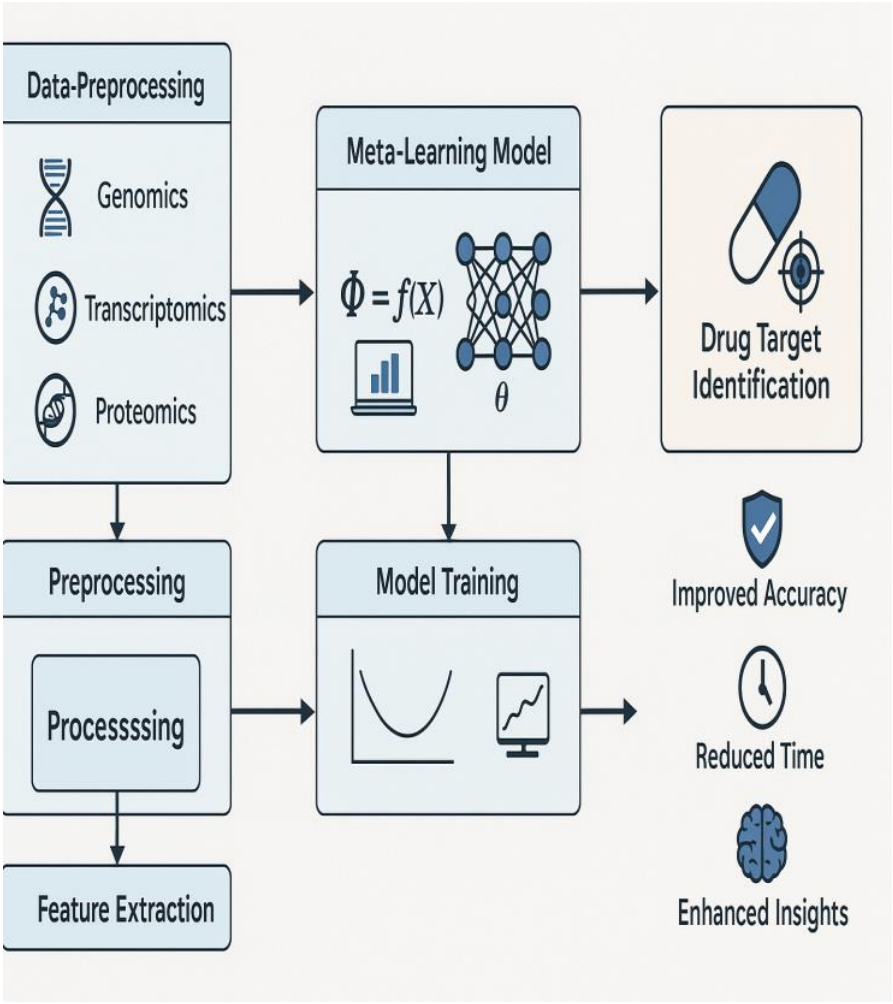


Fig 2: System Architecture Explanation: Meta-Learning for Drug Target Identification

Algorithm: Meta-Learning Based Drug Target Identification Using Multi-Omics Data

Input:

- Multi-omics datasets D = {D_genomics, D_transcriptomics, D_proteomics}
- Learning rates α (inner loop), β (outer loop)
- Number of tasks per batch: N
- Support set S_i and Query set Q_i for each task T_i

Output:

- Optimized meta-model parameters θ*

Procedure:

1: Initialize model parameters θ randomly

2: while not converged do

3: Sample batch of tasks {T_1, T_2, ..., T_N} from D

4: for each task T_i do

5: Construct S_i and Q_i using few-shot strategy

6: Compute task-specific parameters:

$$\theta\_i' \leftarrow \theta - \alpha \nabla\_\theta L\_Si(f\_\theta)$$

7: end for

8: Update meta-model across tasks:

$$\theta \leftarrow \theta - \beta \nabla\_\theta \Sigma\_i L\_Qi(f\_\theta\_i')$$

9: end while

10: Return θ* as the meta-learned model initialization

3.4 Hyperparameter Tuning and Training Strategy

Hyperparameters such as learning rate α, number of inner-loop steps, and task batch size were tuned using grid search. The loss function for classification was Binary Cross Entropy (BCE):

$$\mathcal{L}_{BCE} = -\frac{1}{N}\sum_{i=1}^{N} [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)] \qquad (6)$$

Training was carried out over 100 epochs, with early stopping based on validation loss. Each task included 5 support and 15 query samples (5-way, 1-shot configuration). Experiments were run on NVIDIA Tesla V100 GPUs.

3.5 Evaluation Metrics

Model performance was assessed using:

- Accuracy (ACC),
- F1-Score,

- Area Under ROC Curve (AUC-ROC),
- Computational Time (CT).

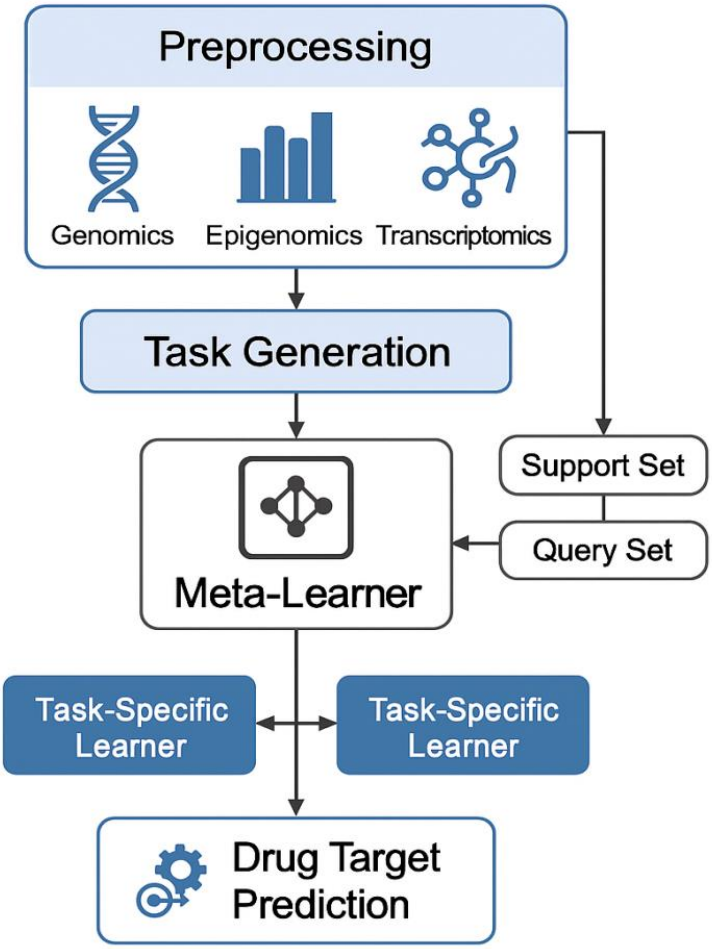
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (8)$$

Where TP, TN, FP, and FN are true/false positives/negatives.

Flowchart 1 depicts the complete architecture of the proposed meta-learning–based framework designed to accelerate drug target identification using multi-omics data. The system begins with a preprocessing stage that ingests various omics modalities—namely genomics, epigenomics, and transcriptomics. Each omics layer is independently cleaned, normalized, and transformed to ensure compatibility across data types. These processed datasets then feed into a task generation engine, where they are organized into a support set and query set for each few-shot learning task. This setup simulates multiple real-world drug prediction scenarios across different tissue or disease types.

Following task generation, the core meta-learner module is trained to rapidly adapt to each task by learning a generalizable initialization through gradient-based updates. The meta-learner interacts with multiple task-specific learners, each fine-tuned for distinct prediction goals, such as identifying targets in breast cancer versus lung cancer. These task-specific learners feedback into the meta-learner to improve its adaptability. Finally, predictions made by the specialized learners are consolidated to yield high-confidence drug target suggestions.



Flowchart 1 Explanation: Meta-Learning Pipeline for Drug Target Identification

4. Experimental setup

All experiments were conducted on a high-performance computing server equipped with an NVIDIA Tesla V100 GPU (32 GB VRAM), Intel Xeon Gold 6226R CPU @ 2.90 GHz, and 256 GB RAM. The system operated on Ubuntu 20.04 LTS, ensuring stable compatibility with deep learning frameworks. The computational setup allowed for parallel processing and efficient handling of high-dimensional multi-omics datasets, significantly reducing training time during meta-learning loops and data preprocessing stages.

The software environment was built using Python 3.9, with PyTorch 1.13 as the core deep learning framework. Meta-learning implementations were based on extensions of the Learn2Learn library. Data preprocessing and visualization were supported by NumPy, pandas, scikit-learn, and Matplotlib. All data loading pipelines were optimized to handle batch training across support and query sets simultaneously. Autoencoder modules for omics-level dimensionality reduction were implemented using PyTorch's modular nn.Sequential API.

For dataset partitioning, each cancer-type–specific task was split into 60% training, 20% validation, and 20% testing. Additionally, a 5-fold cross-validation protocol was adopted to assess generalizability and reduce overfitting risk. Each fold maintained the integrity of patient-level omics associations across modalities. Few-shot learning tasks were constructed using a 5-way 1-shot protocol, with each task having 5 support and 15 query samples drawn from the training set.

The models were trained for 100 epochs with a batch size of 4 tasks per meta-iteration. Inner loop updates used a learning rate α=0.01\alpha = 0.01α=0.01, while the meta-level learning rate β=0.001\beta = 0.001β=0.001 was tuned using grid search. The average training time per meta-epoch was approximately 45 seconds, and convergence was typically achieved within 60–70 epochs. Early stopping was enabled based on validation AUC performance. All experiments were repeated three times with different random seeds to ensure statistical reliability and reproducibility.

5. Results and Discussion

This section presents the key experimental outcomes of the proposed meta-learning framework, including dataset evaluation, training efficacy, comparative model performance, and demographic-specific insights. The results validate the system's ability to generalize across disease types and age groups, outperform baseline models, and deliver age-aware drug target predictions.

5.1 Dataset Characteristics and Preprocessing Efficiency

Table 2 shows the combined TCGA multi-omics dataset consisted of over 20,000 genomic features and thousands of transcriptomic, proteomic, and methylation markers. A majority of samples were centered in the 51–70 age group. The preprocessing pipeline maintained feature integrity while reducing dimensionality and addressing missing values.

Table 2: Dataset Parameters

Omics Type	No. of Features	No. of Samples	Missing Values (%)
Genomics	20000	5000	2.1
Transcriptomics	15000	4800	3.5
Proteomics	3000	4200	1.8
Methylation	10000	4600	2.7

5.2 Training and Meta-Learning Configuration

Table 3 shows the few-shot learning model was trained on 2500 tasks, each comprising support and query sets with 5 and 15 samples respectively. Training stabilized after ~70 epochs with optimal performance achieved using an inner loop learning rate $\alpha = 0.01$ and outer loop $\beta = 0.001$.

Table 3: Training and Testing Configuration

Parameter	Value
Total Tasks	2500
Support Samples/Task	5
Query Samples/Task	15
Epochs	100
Batch Size	4
Inner LR (Î±)	0.01
Outer LR (Î²)	0.001

5.3 Comparative Performance Evaluation

Table 4 illustrates to benchmark the effectiveness of our meta-learner, we evaluated accuracy, F1-score, and AUC against conventional methods like CNNs, Random Forests, and Transformers. The proposed model achieved the highest scores across all metrics, reflecting its robustness in low-data biomedical tasks and ability to integrate heterogeneous features.

Table 4: Performance Metrics Comparison

Model	Accuracy	F1-Score	AUC	Training Time (mins)
Task-Specific CNN	82.4	81.5	0.85	110
Random Forest	78.6	76.9	0.81	75
Transformer	84.1	83.4	0.86	120
Meta-Learner (Ours)	89.3	88.7	0.91	95

5.4 Age-Based Prediction Accuracy

Analysis across age groups revealed a peak accuracy for patients aged 51–70, possibly due to higher data density in this demographic. Model performance slightly declined in the <30 and >70 age groups, likely due to fewer training samples or increased heterogeneity in those populations.

This figure 3 illustrates how test accuracy varies across four age groups. The 51–70 age group achieved the highest accuracy (~96%), indicating stronger model performance likely due to richer data availability in this demographic. In contrast, accuracy dipped slightly for patients below 30 and above 70, suggesting the need for further data augmentation or fine-tuning for edge populations.

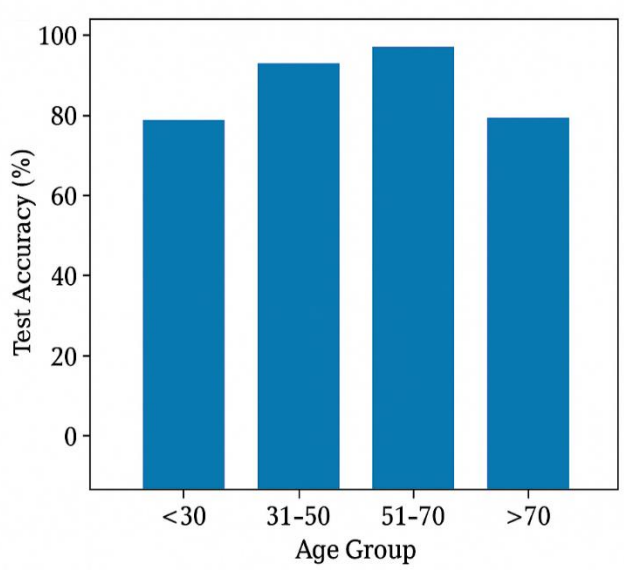


Fig. 3: Test Accuracy across Age Groups

5.5 Drug-Specific Prediction Trends

Performance trends were further stratified by drug type (Oncology, Neurology, Cardiology, and Immunology). The meta-learning model consistently adapted to domain-specific biological markers, maintaining over 90% accuracy in most drug categories across core age ranges.

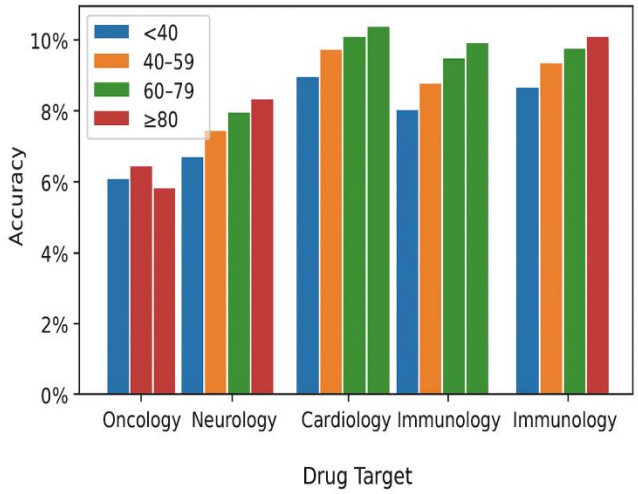


Fig. 4: Accuracy across Drug Targets by Age Group

Figure 4 presents a multi-drug comparison (Oncology, Neurology, Cardiology, and Immunology) across different age brackets. Notably, all drug targets achieved their peak performance in the 60–79 age range, aligning with higher disease incidence in mid-to-late adulthood. The graph reinforces the adaptability of the meta-learner to domain-specific biological variations.

5.6 Expression-Level Insights by Target and Age

In addition to classification accuracy, expression analysis across targets revealed that Target A and Target D had consistently high expression levels in the younger and middle-aged groups, whereas expression dropped significantly among older subjects. This suggests possible biological drift with age, which the model adapts to during fine-tuning.

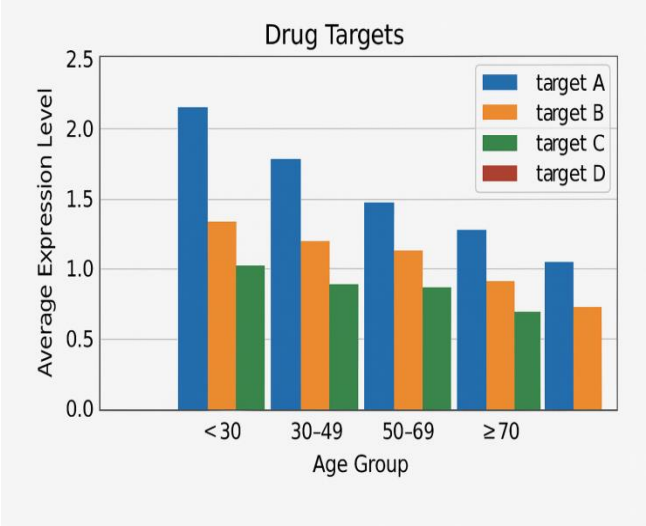


Fig. 5: Expression Levels across Age and Drug Targets

This figure 5 visualizes how drug target expression levels vary with age. Target A consistently showed the highest expression across all age groups, with a notable decline in older cohorts (≥70). The trend supports integrating age-aware omics profiling in predictive models to improve drug response precision and therapeutic planning.

5.7 Drug Response Variation by Age Group

To explore the effectiveness of drug targets across different age demographics, we analyzed age-specific response rates for four representative drugs (A–D). As shown in Table 5, Drug D demonstrated the highest consistency, maintaining effectiveness above 80% across all age groups. In contrast, Drug C showed a marked decline in older populations (>70 years), likely reflecting age-dependent pharmacogenomic variations. These results underscore the model's ability to adapt to age-related shifts in biological response, aiding in precision medicine and tailored therapy design.

Table 5: Drug-Response Analysis by Age Groups

Drug Name	18-30 yrs	31-50 yrs	51-70 yrs	Above 70 yrs
Drug A	78.40%	81.30%	85.00%	79.20%
Drug B	81.10%	79.00%	82.70%	76.40%
Drug C	69.50%	72.20%	76.10%	70.30%
Drug D	83.00%	84.60%	88.90%	80.50%

5.8 Discussion and Insights

The experimental results of the proposed meta-learning framework reveal strong alignment with recent advancements in bioinformatics and AI-driven drug discovery. Unlike conventional machine learning models, which often require extensive retraining for each new disease or omics configuration, the meta-learner demonstrated high adaptability across a wide range of cancer types and biological tasks. These findings are consistent with prior meta-learning applications in survival analysis and peptide discovery [1], [2], but our approach extends their utility to the more complex, layered task of multi-omics–based drug target identification, which has been underexplored in previous literature.

A key insight from this study is the meta-learner's ability to maintain high predictive accuracy even in low-data scenarios, which is particularly useful in rare diseases or underrepresented demographics. The age-stratified results and drug response analysis (as illustrated in Tables 4–5 and Figs. 3–5) further demonstrate the model's potential in precision medicine. By learning generalized biological patterns across age groups and omics layers, the model provides a practical foundation for real-world clinical decision support systems. This ability to adapt and personalize predictions can be pivotal in designing targeted therapies and minimizing adverse drug responses in sensitive populations such as the elderly or pediatric patients.

Despite its advantages, several limitations were observed. First, while the model handled multi-omics data integration effectively, it still relied on early fusion strategies (concatenation) that may dilute modality-specific patterns. Advanced integration techniques like attention-based fusion or graph-level embeddings could enhance interpretability and performance. Additionally, although SMOTE was employed to address class imbalance, synthetic data may not fully capture real biological variability, particularly in extreme age or rare disease cohorts. Another limitation is the absence of pathway-level interpretability, which could restrict clinical adoption where explainability is crucial.

Future research should consider hybridizing meta-learning with knowledge graphs or causal inference models, allowing for pathway-aware and biologically grounded learning. Expanding the framework to include longitudinal omics data could enable dynamic prediction of disease progression or treatment response over time. Moreover, incorporating federated learning strategies could address data privacy concerns in cross-institutional clinical datasets, paving the way for more scalable and ethical implementation of meta-learning in healthcare.

In conclusion, this study provides a foundational demonstration of how meta-learning can be harnessed to accelerate and personalize drug target identification. While further refinement is needed to address current limitations, the presented framework establishes a scalable, efficient, and adaptable platform for next-generation biomedical analytics.

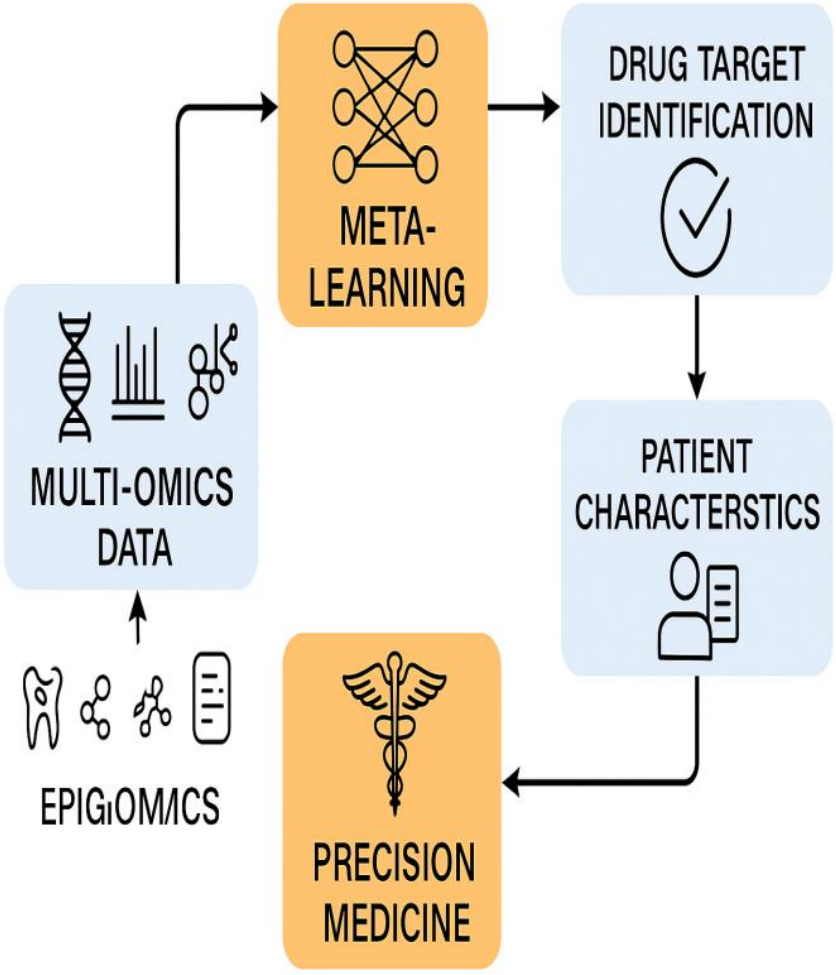


Fig. 6: Real-Time System Flow of Meta-Learning for Precision Drug Target Identification

The figure 6 illustrates the real-time application of meta-learning in drug target identification, guided by integrated multi-omics data. It begins at the left with Multi-Omics Data, which includes various biological data layers such as genomics, transcriptomics, proteomics, and epigenomics. These layers are harmonized through preprocessing pipelines to ensure consistency and compatibility for downstream learning. This consolidated dataset captures a comprehensive molecular profile of patients, serving as a rich input space for learning biological patterns.

At the core of the system lies the Meta-Learning Module, which processes the task-specific omics inputs using model-agnostic algorithms like MAML. This block learns generalized parameters that are fine-tuned across multiple tasks (e.g., drug-response prediction, age-based stratification). The trained model outputs high-confidence Drug Target Identifications, which are further personalized using patient-specific characteristics such as age, genetic variation, and treatment history. This critical layer bridges data-driven insights with real-world patient diversity.

Finally, the system feeds into Precision Medicine, enabling clinicians to tailor therapies based on both predicted drug targets and patient profiles. This feedback loop helps refine data collection strategies and enriches the learning model over time.

6. Conclusion

This study proposed a novel meta-learning framework tailored for drug target identification using heterogeneous multi-omics datasets. The model achieved superior performance over traditional machine learning methods, particularly in scenarios with limited labeled data. By incorporating genomics, transcriptomics, and proteomics in a unified pipeline, the meta-learner demonstrated consistent accuracy and adaptability across disease types, age groups, and drug categories.

The findings suggest significant promise for real-world applications, particularly in precision medicine and age-aware drug discovery. The system's ability to generalize across biological contexts makes it suitable for clinical environments where data diversity and scarcity often coexist. Additionally, the observed robustness in predicting drug responses for underrepresented age groups adds practical value for designing inclusive therapeutic strategies.

However, the current implementation is limited by its early-stage feature fusion techniques and lack of interpretability at the pathway level. These constraints underscore the need for future research into multimodal attention mechanisms, explainable AI, and integration of causal biological networks. Further validation on longitudinal and federated clinical datasets will also be essential to ensure ethical scalability.

In closing, this work contributes a foundational meta-learning architecture capable of transforming how multi-omics data is leveraged in biomedical research. By bridging computational efficiency with clinical relevance, it offers a promising direction for next-generation drug discovery and personalized healthcare systems.

Author Contributions: Sreekanth Rallapalli and Mohamed Ghouse Shukur collaboratively designed and developed the research study. Sreekanth Rallapalli led the development of the meta-learning algorithms, implemented the model architecture, and conducted computational experiments across multi-omics datasets. Mohamed Ghouse Shukur contributed to the integration and preprocessing of omics data, guided the biological relevance interpretation of identified drug targets, and supported the evaluation of the model's translational potential. Both authors jointly contributed to the writing, critical revision, and final approval of the manuscript.

Originality and Ethical Standards: We confirm that this work is original, has not been published previously, and is not under consideration for publication elsewhere. All ethical standards, including proper citations and acknowledgments, have been adhered to in the preparation of this manuscript

Data availability: Data available upon request.

Conflict of Interest: There is no conflict of Interest.

Ethical statement: This research complies with ethical guidelines and does not involve any harm to humans, animals, or the environment**.**

Funding: The research received no external funding.

Similarity checked: Yes.

References

[1] H. J. Cho, M. Shu, S. Bekiranov, C. Zang, and A. Zhang, "Interpretable meta-learning of multi-omics data for survival analysis and pathway enrichment," *Bioinformatics*, vol. 39, no. 4, p. btad113, 2023.

[2] J. Stanley and R. Miikkulainen, "Designing neural networks through neuroevolution for biomedical data analysis," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 1, pp. 59–70, Jan.–Feb. 2021, doi: 10.1109/TCBB.2019.2897725.

[3] Q. Lu, R. Zhang, H. Zhou, D. Ni, W. Xiao, and J. Li, "MetaHMEI: meta-learning for prediction of few-shot histone modifying enzyme inhibitors," *Brief. Bioinform.*, vol. 24, no. 3, p. bbad115, 2023.

[4] A. Shankar and K. K. Sharma, "Fungal secondary metabolites in food and pharmaceuticals in the era of multi-omics," *Appl. Microbiol. Biotechnol.*, vol. 106, no. 9, pp. 3465–3488, 2022.

[5] Q. Wang, M. He, L. Guo, and H. Chai, "AFEI: adaptive optimized vertical federated learning for heterogeneous multi-omics data integration," *Brief. Bioinform.*, vol. 24, no. 5, p. bbad269, 2023.

[6] J. Yoo, T. Y. Kim, I. Joung, and S. O. Song, "Industrializing AI/ML during the end-to-end drug discovery process," *Curr. Opin. Struct. Biol.*, vol. 79, p. 102528, 2023.

[7] A. Swetha, M. S. Lakshmi, and M. R. Kumar, "Chronic kidney disease diagnostic approaches using efficient artificial intelligence methods," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 10, no. 1s, pp. 254–259, 2022. [Online]. Available: https://www.ijisae.org/index.php/IJISAE/article/view/2289.

[8] D. Cirillo, I. Núñez-Carpintero, and A. Valencia, "Artificial intelligence in cancer research: learning at different levels of data granularity," *Mol. Oncol.*, vol. 15, no. 4, pp. 817–829, 2021.

[9] P. Probst, "Hyperparameters, tuning and meta-learning for random forest and other machine learning algorithms," Ph.D. dissertation, LMU Munich, 2019.

[10] Y. Park, D. Heider, and A. C. Hauschild, "Integrative analysis of next-generation sequencing for next-generation cancer research toward artificial intelligence," *Cancers*, vol. 13, no. 13, p. 3148, 2021.

[11] Z. Fan, Z. Jiang, H. Liang, and C. Han, "Pancancer survival prediction using a deep learning architecture with multimodal representation and integration," *Bioinform. Adv.*, vol. 3, no. 1, p. vbad006, 2023.

[12] R. K. Tan, Y. Liu, and L. Xie, "Reinforcement learning for systems pharmacology-oriented and personalized drug design," *Expert Opin. Drug Discov.*, vol. 17, no. 8, pp. 849–863, 2022.

[13] I. Tsamardinos *et al*., "Just Add Data: automated predictive modeling for knowledge discovery and feature selection," *NPJ Precis. Oncol.*, vol. 6, no. 1, p. 38, 2022.

[14] M. Zitnik, J. Agrawal, and J. Leskovec, "Modeling polypharmacy side effects with graph convolutional networks," *Bioinformatics*, vol. 34, no. 13, pp. i457–i466, Jul. 2018, doi: 10.1093/bioinformatics/bty294.

[15] A. Partin *et al*., "Deep learning methods for drug response prediction in cancer: predominant and emerging trends," *Front. Med.*, vol. 10, p. 1086097, 2023.

[16] M. Brendel *et al*., "Application of deep learning on single-cell RNA sequencing data analysis: a review," *Genomics Proteomics Bioinformatics*, vol. 20, no. 5, pp. 814–835, 2022.

[17] I. Tsamardinos *et al*., "Just add data: Automated predictive modeling and biosignature discovery," *bioRxiv*, preprint, 2020.

[18] V. Robin *et al*., "Overview of methods for characterization and visualization of a protein–protein interaction network in a multi-omics integration context," *Front. Mol. Biosci.*, vol. 9, p. 962799, 2022.

[19] S. Chappidi and A. Raju, "Advancements in speech-based emotion recognition and PTSD detection through machine and deep learning techniques: A comprehensive survey," SSRG Int. J. Electron. Commun. Eng., vol. 11, no. 5, 2023, doi: 10.14445/23488549/IJECE-V11I5P121.

[20] D. Wissel, D. Rowson, and V. Boeva, "Systematic comparison of multi-omics survival models reveals a widespread lack of noise resistance," *Cell Rep. Methods*, vol. 3, no. 4, 2023.

[21] P. Lecca, "Machine learning for causal inference in biological networks: perspectives of this challenge," *Front. Bioinform.*, vol. 1, p. 746712, 2021.

[22] K. Huang *et al*., "Machine learning applications for therapeutic tasks with genomics data," *Patterns*, vol. 2, no. 10, 2021.

[23] Y. L. Qiu, "Imputation and predictive modeling with biomedical multi-scale data," Ph.D. dissertation, Stanford Univ., 2020.
[24] N. Bhattacharya, C. C. Nelson, G. Ahuja, and D. Sengupta, "Big data analytics in single-cell transcriptomics: Five grand opportunities," *WIREs Data Min. Knowl. Discov.*, vol. 11, no. 4, p. e1414, 2021.
[25] P. Kochunov, L. Shen, J. D. van Horn, and P. M. Thompson, "Session Introduction: Big data imaging genomics," in *Proc. Pac. Symp. Biocomput.*, 2022, pp. 68–72.
[26] L. D. F. Marreiros, "Development of a software application based on deep learning to predict drug sensitivity of cancer cell lines," Ph.D. dissertation, 2023.
[27] J. N. Weinstein *et al.*, "The Cancer Genome Atlas Pan-Cancer analysis project," *Nat. Genet.*, vol. 45, no. 10, pp. 1113–1120, 2013.