Few-Shot Learning for Multi-Omics Disease Classification with a MAML-Based Model

Ganghui YI [a1], Jeongjin JU [a1], Kyuri JO [a2]
[a] *Department of Computer Engineering, Chungbuk National University, Cheongju, Republic of Korea*

Abstract. Few-shot learning offers a promising approach for disease classification in settings where labeled data are scarce. While widely explored in cancer research, its application to non-cancer diseases and multi-omics data remains limited. In this study, we propose a MAML-based few-shot learning model, pre-trained on TCGA data from four distinct tissue types. We then evaluate its adaptability across three disease categories, including COVID-19, Cirrhosis, and HBV-HCC. Our results demonstrate that MAML consistently outperforms a baseline MLP, achieving higher PR-AUC and ROC-AUC for COVID-19 and Cirrhosis. However, for HBV-HCC, where disease characteristics closely align with the pre-training data, the baseline MLP exhibits slightly superior performance. These findings highlight MAML's potential for low-data disease classification, while also underscoring conditions where its benefits may be limited.

Keywords. Few-shot learning, MAML, Meta-learning, Disease Classification, Multi-omics

1. Introduction

Disease classification is a cornerstone of medical research and clinical diagnostics, yet it often encounters the challenge of limited labeled data. Traditional deep learning approaches require extensive, high-quality annotations for robust performance, but these are often unavailable in medical settings due to cost, privacy issues, and the difficulty of obtaining diverse patient samples. Consequently, there is a growing need for learning frameworks that can perform effectively with very few training examples.

Recent advances in few-shot learning offer a promising solution by enabling models to generalize from minimal supervision. While few-shot learning has been widely explored in cancer research [1,2,3], its application to non-cancer diseases remains underexplored. Additionally, most prior studies focus on image-based medical tasks [3], whereas applications in multi-omics and gene expression-based classification, which are crucial for precision medicine, have received far less attention. Beyond cancer, few-shot learning has been successfully applied to other biomedical problems, such as protein structure

[1]These authors contributed equally.
[2]Corresponding author.

prediction and drug discovery [4,5], demonstrating its potential across a variety of medical domains. However, many existing few-shot learning approaches in biomedical applications have been designed with task-specific optimizations, which present challenges when generalizing to diverse disease datasets.

Among few-shot learning techniques, Model-Agnostic Meta-Learning (MAML) has gained prominence due to its ability to learn a generalizable initialization that enables rapid adaptation to new tasks with only a few gradient updates [6]. While conventional supervised learning optimizes a model for a specific task, MAML is meta-trained across multiple tasks. As a result, it learns an initialization that can be quickly fine-tuned to new diseases with minimal labeled data. This makes MAML particularly well-suited for biomedical applications, where the continual emergence of new diseases sustains the demand for flexible and rapidly adaptable classification methods, even when labeled data are scarce.

To bridge the gap in few-shot learning research for multi-omics classification beyond cancer, we propose a MAML-based few-shot learning approach that extends its applicability to a broader range of diseases. Our study leverages The Cancer Genome Atlas (TCGA) datasets for meta-training and evaluates the model's adaptability in few-shot settings across both cancerous and non-cancerous diseases. Unlike previous applications of few-shot learning in medical AI, which have primarily focused on image-based tasks or single-disease classification, our approach applies MAML to multi-omics data, allowing it to generalize across a broad spectrum of diseases. Through these experiments, we aim to validate our hypothesis that a MAML-based model outperforms a conventional machine learning model in few-shot scenarios, particularly in non-cancer diseases where labeled data are scarce. Our findings highlight the potential of MAML-based meta-learning for biomedical classification and demonstrate its generalizability in low-data disease classification tasks.

2. Method

In this study, we proposed a few-shot learning framework for disease prediction that integrated pre-training, meta-training, and meta-testing. To effectively leverage genomic data, our approach consisted of four key stages: dataset collection from publicly available repositories; systematic data preprocessing to ensure consistency, address class imbalance, and reduce dimensionality via gene filtering; model design based on a MAML architecture with weight initialization using a pre-trained Multi-Layer Perceptron (MLP); and comprehensive performance evaluation. Figure 1 provides an overview of the experimental design.

2.1. Dataset

We utilized two types of omics data, mRNA and miRNA, obtained from high-throughput sequencing in both the meta-training and meta-testing phases. TCGA datasets used for meta-training were sourced from Ron Shamir's group [7]. We focused on cancers associated with four tissue types: breast (Breast Invasive Carcinoma, BRCA), kidney (Kidney Renal Clear Cell Carcinoma, KIRC), liver (Liver Hepatocellular Carcinoma, LIHC), and lung (Lung Squamous Cell Carcinoma, LUSC). For the meta-testing phase, diseases

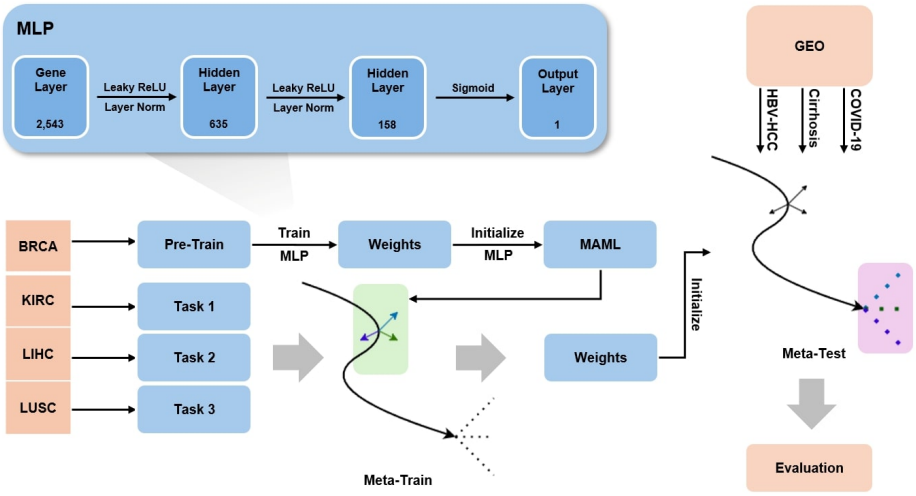

Figure 1. Schematic overview of the experimental workflow, illustrating the process of dataset collection, pre-training an MLP for weight initialization, and applying MAML for meta-training and meta-testing. The top section shows the shared MLP architecture used for both pre-training and classification in MAML and baseline models.

(i.e., prediction targets) were filtered based on their disease-disease associations with all four TCGA datasets, computed using the Jaccard Index (JI) from the DisGeNet database [8]. To ensure relevance and consistency, we prioritized diseases with the highest JI values and confirmed their association across all selected TCGA tissue types. We utilized three target diseases for our experiment: Coronavirus disease (COVID-19), Cirrhosis, and Hepatitis B Virus-Associated Hepatocellular Carcinoma (HBV-HCC) [9,10,11]. The National Center for Biotechnology Information (NCBI)-generated RNA-seq count data for these samples were obtained from the Gene Expression Omnibus (GEO) repository. Finally, we used biological pathways from the Reactome database to refine our gene selection and reduce dimensionality [12]. Only genes associated with the "Disease" and "Signal Transduction" pathways were retained to focus on biologically relevant features. Table 1 summarizes both the TCGA and GEO datasets, including the number of healthy controls (Label 0) and patients (Label 1), along with the total number of samples. Note that for the TCGA datasets, we downsampled the positive class to achieve a 3:1 ratio of positive to negative.

2.2. Preprocessing

All mRNA and miRNA gene symbols were unified according to the conventions of the TCGA datasets provided by Ron Shamir's group: mRNA genes adhered to standard gene nomenclature, while miRNAs were labeled with the "hsa-" prefix. Additionally, all mRNA and miRNA data used in both the meta-training and meta-testing phases underwent log transformation followed by standardization to stabilize variance and ensure comparability across features. For the TCGA datasets prepared for meta-training, a class imbalance was identified, with a positive-to-negative class ratio of approximately

Table 1. Summary of Meta-Training (TCGA) and Meta-Testing (GEO) Datasets.
Note: The TCGA data were downsampled to a 3:1 (Label 1 : Label 0) ratio. Label 0 indicates healthy controls and Label 1 indicates patient samples.

Data Source	Cancer/Disease Type	Label 0 (Healthy)	Label 1 (Patient)	Total Samples
TCGA	BRCA	87	261	348
	KIRC	71	213	284
	LIHC	50	150	200
	LUSC	38	114	152
GEO	COVID-19	4	6	10
	Cirrhosis	24	150	174
	HBV-HCC	21	21	42

10:1. To address this, the positive class was downsampled, resulting in a final ratio of around 3:1. Lastly, gene filtering was performed by intersecting the gene sets from the TCGA and Reactome databases. The filtered genes were further intersected with NCBI-generated GEO data to ensure consistency, resulting in a final set of 2,543 genes. The expression levels of these transcripts were subsequently used as input features for our MAML-based few-shot learning framework, which is elaborated upon in the following section on Model Design.

2.3. Model Design

We employed a MAML-based few-shot learning framework with weight initialization derived from pre-training. The architecture for both the MLP used in pre-training and the MAML model consisted of an input layer with 2,543 features, two hidden layers of 635 and 158 neurons using Leaky ReLU activations, and an output layer for binary classification. To provide a meaningful initialization for MAML, the MLP was pre-trained on the TCGA breast cancer dataset, which had the largest sample size among the four tissue datasets [13]. The weights obtained from this pre-training phase were used to initialize the parameters for meta-training in the MAML framework.

During meta-training, the MAML model was trained on TCGA datasets from three tissue types: kidney, liver, and lung. A 2-way k-shot learning setup was adopted, with $k = 1, 3, 5$. To enable shallow generalization, a single gradient update was performed per task using an inner loop learning rate (α) of 0.001. After task-specific adaptation, we employed AdamW [14], an Adam optimizer variant that decouples weight decay (L_2 regularization) from the adaptive gradient updates, to perform the meta-update step using an outer loop learning rate (β) of 0.001. AdamW was chosen to stabilize optimization by preventing excessively large weight updates, thereby improving generalization across tasks. Given the small meta-batch size ($n = 3$), there was an increased risk of overfitting to a specific tissue. To mitigate this, we implemented an early stopping criterion to halt training when the average loss across meta-batches no longer decreased. Specifically, training was terminated if the meta-loss did not improve for 100 consecutive outer-loop iterations (patience = 100). Furthermore, Layer Normalization was introduced instead of Batch Normalization to ensure stable training under these small-batch conditions [15, 16]. For meta-testing, the inner loop learning rate (α) was increased to 0.01 to facilitate faster task-specific adaptation, and each task underwent 10 gradient updates to enable

deeper task-specific learning and enhance predictive performance. The model was then evaluated on three distinct target diseases described in the Dataset section. A formal description of the model's training and adaptation process is outlined in Algorithm 1.

Algorithm 1 MAML for Few-Shot Multi-Omics Disease Classification

Require: $\mathcal{D}_{\text{pre}}$: Pre-training dataset (BRCA)
Require: $p(\mathcal{T}_{\text{train}})$: Meta-training task distribution (KIRC, LIHC, LUSC)
Require: $\mathcal{D}_{\text{test}}$: Meta-testing datasets (COVID-19, Cirrhosis, HBV-HCC)
Require: α_{train}, α_{test}: Inner-loop learning rates for meta-training and meta-testing
Require: β: Meta-update (outer-loop) learning rate
Require: k: Number of support samples per task
Require: n_{test}: Number of inner-loop updates during meta-testing
Pre-training:
  Train an MLP on $\mathcal{D}_{\text{pre}}$ to obtain initial parameters θ_0
  Initialize $\theta \leftarrow \theta_0$
Meta-Training:
while not converged **do**
  Sample a batch of tasks $\{\mathcal{T}_i\} \sim p(\mathcal{T}_{\text{train}})$
  for each task $\mathcal{T}_i$ in the batch **do**
    Sample k per class for support set $\mathcal{D}_i$ from $\mathcal{T}_i$
    Compute gradient: $g_i = \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(\theta; \mathcal{D}_i)$
    Update task-specific parameters: $\theta_i' \leftarrow \theta - \alpha_{\text{train}} \cdot g_i$
    Sample query set $\mathcal{D}_i'$ from $\mathcal{T}_i$
  end for
  Meta-update: $\theta \leftarrow \theta - \beta \sum_i \nabla_{\theta_i'} \mathcal{L}_{\mathcal{T}_i}(\theta_i'; \mathcal{D}_i')$
  if early stopping criterion satisfied **then**
    break
  end if
end while
Meta-Testing:
for each new task $\tilde{T}$ in $\mathcal{D}_{\text{test}}$ **do**
  Split $\tilde{T}$ into support set $\mathcal{S}_{\tilde{T}}$ and query set $\mathcal{Q}_{\tilde{T}}$
  Initialize task parameters: $\theta_{\tilde{T}} \leftarrow \theta$
  for $j = 1$ to n_{test} **do**
    Compute gradient: $g_{\tilde{T}} = \nabla_\theta \mathcal{L}_{\tilde{T}}(\theta_{\tilde{T}}; \mathcal{S}_{\tilde{T}})$
    Update parameters: $\theta_{\tilde{T}} \leftarrow \theta_{\tilde{T}} - \alpha_{\text{test}} \cdot g_{\tilde{T}}$
  end for
  Evaluate performance of $\theta_{\tilde{T}}$ on $\mathcal{Q}_{\tilde{T}}$
end for

2.4. Evaluation Metric

The performance of disease prediction was evaluated using the Precision-Recall Area Under the Curve (PR-AUC) and Receiver Operating Characteristic Area Under the Curve (ROC-AUC). Given that the positive class holds greater importance in the context of disease prediction, PR-AUC was used as the primary metric due to its focus on precision

and recall. ROC-AUC was additionally reported to provide a comprehensive evaluation across both classes. Both metrics are threshold independent, making them well-suited for direct model comparisons with baseline models and robust performance evaluation [17].

3. Result and Discussion

Table 2 presents the performance comparison between the baseline MLP and our proposed MAML-based approach under three k-shot settings for each target disease. All results are reported as the mean and standard deviation computed over 50 meta-testing tasks. Overall, the proposed model shows consistently higher performance for COVID-19 and Cirrhosis across PR-AUC and ROC-AUC. For COVID-19, MAML exceeds the baseline by an average of 0.09 in PR-AUC and 0.13 in ROC-AUC, while in Cirrhosis, it achieves an average improvement of 0.09 in PR-AUC and 0.11 in ROC-AUC. In contrast, for HBV-HCC, both models exhibit exceptionally high performance, with the baseline MLP showing slightly better results by approximately 0.02 in PR-AUC and 0.03 in ROC-AUC.

Table 2. Performance Comparison Between Baseline (MLP) and Proposed (MAML) Models.
Note: While the proposed MAML model uses a 2-way k-shot setting, the baseline MLP model is trained on $2 \times k$ samples for each task, as it does not follow the n-way k-shot paradigm.

Target Disease	Metric	k-shot	Baseline (MLP)	Proposed (MAML)
COVID-19	PR-AUC	1	0.69 (±0.12)	0.75 (±0.08)
		3	0.66 (±0.18)	0.77 (±0.15)
		5	N/A	N/A
	ROC-AUC	1	0.48 (±0.13)	0.56 (±0.08)
		3	0.24 (±0.30)	0.42 (±0.29)
		5	N/A	N/A
Cirrhosis	PR-AUC	1	0.61 (±0.13)	0.61 (±0.09)
		3	0.63 (±0.14)	0.80 (±0.14)
		5	0.68 (±0.15)	0.79 (±0.13)
	ROC-AUC	1	0.57 (±0.18)	0.61 (±0.11)
		3	0.61 (±0.17)	0.78 (±0.13)
		5	0.66 (±0.17)	0.78 (±0.12)
HBV-HCC	PR-AUC	1	0.996 (±0.01)	0.977 (±0.02)
		3	1.000 (±0.00)	0.983 (±0.01)
		5	1.000 (±0.00)	0.982 (±0.01)
	ROC-AUC	1	0.996 (±0.01)	0.955 (±0.03)
		3	1.000 (±0.00)	0.970 (±0.03)
		5	1.000 (±0.00)	0.970 (±0.02)

The MAML-based model's consistent advantage in COVID-19 and Cirrhosis classification can be primarily attributed to its meta-learning capability. By leveraging pretrained weights and undergoing meta-training, the model learns a parameter initialization that facilitates rapid adaptation to new tasks, even within very limited training samples. This rapid adaptability is particularly beneficial in medical applications where collect-

ing extensive annotated data can be challenging. In HBV-HCC task, the baseline MLP slightly outperforms the MAML-based model. One plausible explanation is the close alignment between the TCGA cancer data and HBV-HCC. Since the MLP was trained in a single supervised pass on all four TCGA datasets, it acquired broad cancer-specific features closely matching the HBV-HCC profile. In this setting, MAML's meta-learning initialization, designed to enable rapid adaptation for novel or diverse tasks, offers less additional benefit as the baseline is already near-ceiling in terms of cancer-relevant representations.

Initially, we hypothesized that a MAML-based approach would outperform a conventional baseline MLP under few-shot conditions. The results largely validate this hypothesis for COVID-19 and Cirrhosis, where the MAML shows superior classification metrics despite very limited data. This aligns with meta-learning principles, which emphasize the advantage of a meta-trained initialization in low-data scenarios. However, for HBV-HCC, which closely aligns with the TCGA dataset, the additional benefits that MAML typically provides are less pronounced. This indicates that while MAML is effective for diverse tasks and data scarcity, its advantage may diminish when a well-trained baseline already reaches near-ceiling performance.

4. Conclusion

This study demonstrated the effectiveness of a MAML-based few-shot learning approach for disease classification using multi-omics data. Initialized with TCGA datasets, our proposed model achieved superior performance in COVID-19 and Cirrhosis classification compared to the baseline MLP, particularly in low-data settings. This result highlights MAML's ability to generalize across diverse disease tasks with limited labeled data. However, in HBV-HCC classification, where the disease characteristics closely resembled the training data, the baseline MLP performed slightly better. This observation suggests that MAML's advantage is most prominent when the target disease differs significantly from the meta-training data and diminishes when domain similarity is high. Overall, these findings emphasize the importance of both dataset quality and domain alignment in determining few-shot learning performance.

Despite the promising results on three target diseases, its generalizability may be constrained by the relatively limited scope of diseases evaluated. Future work will address this limitation by expanding the range of target diseases to include tissues such as the lungs and breasts, which were previously part of the TCGA pre-training data but have not yet been explored as target tasks. This will provide a more comprehensive assessment of the method's robustness and scalability. Moreover, we plan to integrate a pathway layer into the model architecture to explicitly capture gene-pathway associations, thereby enhancing both interpretability and predictive performance. Finally, we will employ explainable AI (xAI) analyses to ensure that our approach not only achieves high accuracy but also yields insights that are meaningful for clinical or translational research contexts. Altogether, these directions will further broaden the applicability of our approach and underscore its potential as a robust few-shot disease classification paradigm.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. RS-2023-00217022; No. RS-2023-00252745).

References

[1] Okimoto LY, Mendonca-Neto R, Nakamura FG, Nakamura EF, Fenyö D, Silva CT. Few-shot genes selection: subset of PAM50 genes for breast cancer subtypes classification. BMC bioinformatics. 2024;25(1):92.

[2] Li T, Shetty S, Kamath A, Jaiswal A, Jiang X, Ding Y, et al. CancerGPT for few shot drug pair synergy prediction using large pretrained language models. NPJ Digital Medicine. 2024;7(1):40.

[3] Pachetti E, Colantonio S. A systematic review of few-shot learning in medical imaging. Artificial intelligence in medicine. 2024:102949.

[4] Zhang J, Liu S, Chen M, Chu H, Wang M, Wang Z, et al. Few-shot learning of accurate folding landscape for protein structure prediction. arXiv preprint arXiv:220809652. 2022.

[5] Schimunek J, Seidl P, Friedrich L, Kuhn D, Rippmann F, Hochreiter S, et al. Context-enriched molecule representations improve few-shot drug discovery. arXiv preprint arXiv:230509481. 2023.

[6] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: International conference on machine learning. PMLR; 2017. p. 1126-35.

[7] Rappoport N, Shamir R. Multi-omic and multi-view clustering algorithms: review and cancer benchmark. Nucleic acids research. 2018;46(20):10546-62.

[8] Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic acids research. 2020;48(D1):D845-55.

[9] Isnard P, Vergnaud P, Garbay S, Jamme M, Eloudzeri M, Karras A, et al. A specific molecular signature in SARS-CoV-2–infected kidney biopsies. JCI insight. 2023;8(5).

[10] Hlady R, Zhao X, El Khoury L, Wagner R, Luna A, Pham K, et al. Epigenetic heterogeneity hotspots in human liver disease progression. Hepatology. 2024:10-1097.

[11] Yoo S, Wang W, Wang Q, Fiel MI, Lee E, Hiotis SP, et al. A pilot systematic genomic comparison of recurrence risks of hepatitis B virus-associated hepatocellular carcinoma with low-and high-degree liver fibrosis. BMC medicine. 2017;15:1-17.

[12] Sinnott JA, Cai T. Pathway aggregation for survival prediction via multiple kernel learning. Statistics in medicine. 2018;37(16):2501-15.

[13] Lin CF, Lee HY. MAster of PuPpets: Model-Agnostic Meta-Learning via Pre-trained Parameters for Natural Language Generation. In: Neurips Workshop on Meta Learning, abs/2110.15943; 2020. .

[14] Loshchilov I. Decoupled weight decay regularization. arXiv preprint arXiv:171105101. 2017.

[15] Lei Ba J, Kiros JR, Hinton GE. Layer normalization. ArXiv e-prints. 2016:arXiv-1607.

[16] Wu Y, He K. Group normalization. In: Proceedings of the European conference on computer vision (ECCV); 2018. p. 3-19.

[17] Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern recognition. 1997;30(7):1145-59.