MINT: Multimodal Integrated Knowledge Transfer to Large Language Models through Preference Optimization with Biomedical Applications

Da Wu [*1] **Zhanliang Wang** [*1 2] **Quan Nguyen** [1 2] **Zhuoran Xu** [1 2] **Kai Wang** [1 2]

Abstract

The scarcity of high-quality multimodal biomedical data limits the effective supervised fine-tuning (SFT) of Large Language Models (LLMs) for specialized tasks. We introduce MINT (Multimodal Integrated kNowledge Transfer), a framework that aligns unimodal decoder-only models with domain-specific patterns from multimodal biomedical data through preference optimization, primarily implemented using the Odds Ratio Preference Optimization (ORPO) framework. MINT leverages upstream multimodal machine learning models to transfer domain expertise to downstream text-only or image-only LLMs, as demonstrated in two applications: (1) Rare genetic disease prediction from texts; (2) Tissue type classification using cell nucleus images. In both cases, MINT-based models outperform those enhanced with alternative approaches such as Supervised Fine-tuning and Retrieval Augmented Generation, even surpassing much larger foundation models in some scenarios. Our study highlights how MINT effectively grafts the classification strengths of encoder-only models into large decoder-only models, enhancing reasoning abilities, and reducing hallucination in biomedical applications.

1. Introduction

Language Models (LLMs) have achieved remarkable advances in natural language understanding and reasoning (Vaswani et al., 2017; Wolf et al., 2020; McKinzie et al.). Recent models with expanded context windows (Dubey et al., 2024; Meta, 2024a;b) have enhanced their potential for biomedical applications (Kim et al., 2024; do Olmo et al., 2024). Despite these advances, adapting LLMs to domain-specific biomedical tasks remains challenging, especially when multimodal data are scarce or specialized clinical reasoning is required.

Supervised fine-tuning (SFT) (Gunel et al., 2020; Chen et al.) is commonly employed for adapting LLMs to specific tasks. While effective for basic linguistic applications (Al-Moslmi et al., 2020; Etzioni et al., 2005; Daiber et al.), SFT faces critical limitations for complex biomedical tasks. First, SFT primarily learns from positive examples only, lacking the ability to explicitly distinguish between correct and incorrect responses. Second, it struggles with structured prediction and logical reasoning tasks (Wu et al., 2024b; Berglund et al., 2023; Zhu et al., 2024) that require nuanced discrimination between similar patterns. Third, SFT often fails to capture the *internal decision patterns* that domain experts use to differentiate between similar cases, especially in data-scarce biomedical domains. These limitations make it difficult for SFT-based models to achieve expert-level performance on specialized tasks like rare disease diagnosis.

To address these challenges, we propose ***M**ultimodal **I**ntegrated k**N**owledge **T**ransfer* (MINT), a framework that aligns LLMs with domain-specific patterns from multimodal biomedical databases through *preference optimization*. The key innovation of MINT lies in its ability to learn the *internal decision patterns* that distinguish correct from incorrect responses by leveraging preferences generated from upstream multimodal models. Unlike SFT, which only learns what is correct, MINT explicitly learns both what is correct and what is incorrect, enabling more nuanced discrimination between similar cases.

MINT bridges the modality gap between multimodal training data and unimodal inference through two key components: (1) an upstream pipeline where a multimodal ML model generates a preference learning dataset with preferred and unfavored responses, and (2) a downstream pipeline where these preference-labeled inputs align LLMs using techniques like Direct Preference Optimization (DPO) (Rafailov et al., 2024) or Odds Ratio Preference Optimization (ORPO) (Hong et al.). This approach enables the LLM to internalize the decision-making patterns of the multimodal expert model, achieving superior performance even with single-modality inputs during inference.

[*]Equal contribution [1]Children's Hospital of Philadelphia, PA, USA [2]University of Pennsylvania, PA, USA. Correspondence to: Kai Wang <wangk@chop.edu>.

Proceedings of the Workshop on Generative AI for Biology at the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025. Copyright 2025 by the author(s).

We demonstrate MINT's effectiveness on two challenging biomedical tasks. For *rare disease prediction*, we leverage GestaltMML (Wu et al., 2024a), a multimodal model trained on facial images and clinical texts. Our MINT-aligned text-only model significantly outperforms SFT (52.99% vs. 37.53% Top-10 accuracy) and RAG (52.99% vs. 6.52%). Detailed analysis reveals that MINT-enhanced models learn to prioritize disease-specific phenotypic patterns that human experts would consider, while avoiding common misdiagnoses. For *tissue type classification*, MINT substantially improves Llama 3.2-Vision-11B-Instruct's ability to distinguish between histologically similar tissues, nearly doubling the Top-5 accuracy (57.58% vs. 32.21%) compared to the base model.

Importantly, MINT-enhanced models maintain their general capabilities while gaining domain-specific expertise. By grafting the classification strengths of encoder models into decoder-based LLMs through preference optimization, MINT enables models to learn the internal decision patterns that distinguish correct from incorrect responses. This approach offers advantages over both SFT, which lacks contrastive learning signals, and RAG, which introduces latency during inference. Our work establishes MINT as a powerful framework for enhancing LLMs' performance on specialized biomedical tasks by learning expert-level discrimination patterns from multimodal data.

2. Related Work

Multimodal Learning in Biomedicine combines information from different modalities (Wu et al., 2024a; Huang et al., 2023), but typically requires multimodal inputs during inference. MINT transfers knowledge from multimodal models to unimodal LLMs, enabling them to benefit from multimodal data even with single-modality inputs.

Preference Optimization for LLMs has evolved from Reinforcement Learning from Human Feedback (RLHF) (Yu et al., 2022; Engstrom et al., 2020; Engstrom et al.) to more direct methods like DPO (Rafailov et al., 2024) and ORPO (Hong et al.). Unlike traditional applications focused on aligning with human preferences, MINT uniquely applies these techniques to transfer domain-specific decision patterns from expert multimodal models to unimodal LLMs.

LLMs for Biomedical Applications include clinical document summarization (Afantenos et al., 2005; Mamykina et al., 2012), patient scheduling (Tripathi et al., 2024; Qiu et al., 2024), and rare disease diagnosis (Kim et al., 2024; do Olmo et al., 2024). While approaches like RAG (Lewis et al., 2020; Wu et al., 2025) enhance LLMs with external knowledge at inference time, MINT embeds domain expertise by learning internal decision patterns during training for more efficient and nuanced inference.

Knowledge Transfer Between Model Architectures often relies on techniques like knowledge distillation (Gunel et al., 2020) that require parallel training of teacher and student models. MINT offers a novel approach using preference optimization to transfer not just knowledge but also decision-making patterns from encoder-based multimodal models to decoder-based LLMs.

3. Preliminaries

Large Language Models (LLMs). The downstream targets of our knowledge transfer approach are autoregressive decoder-only LLMs. Given an input $x \in \mathcal{X}^{\text{uni}}$, an LLM models the conditional probability of the next token y_t given all previous tokens: $p_\theta(y_t \mid x, y_{<t})$. For classification, we convert labels into text format.

Multimodal Machine Learning (MML) Models. The upstream knowledge sources are multimodal models integrating information from different modalities. For rare disease prediction, we use GestaltMML (Wu et al., 2024a), which combines features from facial images with clinical phenotypes. For tissue classification, we use PLIP (Huang et al., 2023), which integrates histopathological images with diagnostic reports.

Supervised Fine-tuning (SFT). Traditional SFT minimizes the negative log-likelihood loss:

$$\mathcal{L}_{\text{SFT}} = -\log p_\theta(y \mid x), \tag{1}$$

where $p_\theta(y \mid x)$ is the probability of generating the correct label y given input x.

Retrieval-Augmented Generation (RAG) RAG (Lewis et al., 2020; Gao et al., 2023) combines retrieval systems with LLMs to enhance text generation by augmenting the model's input with relevant information retrieved from external knowledge sources. This approach has proven effective for knowledge-intensive tasks across various domains, including question answering and biomedical text processing (Jiang et al., 2023; Wu et al., 2025). RAG enables models to access domain-specific knowledge without requiring extensive fine-tuning, making it particularly valuable for specialized applications.

Preference Optimization. These techniques leverage both preferred responses y^w and non-preferred responses y^l to better align models with desired outputs.

Direct Preference Optimization (DPO) (Rafailov et al., 2024) maximizes the log odds ratio between preferred and non-preferred responses:

$$\mathcal{L}_{\text{DPO}} = -\log \sigma \left(\beta \log \frac{p_\theta(y^w \mid x)}{p_{\text{ref}}(y^w \mid x)} - \beta \log \frac{p_\theta(y^l \mid x)}{p_{\text{ref}}(y^l \mid x)} \right), \tag{2}$$

where p_{ref} is a frozen reference model and β controls regularization strength.

Odds Ratio Preference Optimization (ORPO) (Hong et al.), our primary approach, combines SFT and preference learning:

$$\mathcal{L}_{\text{ORPO}} = \mathcal{L}_{\text{SFT}} + \alpha \cdot \mathcal{L}_{\text{OR}}, \tag{3}$$

where

$$\mathcal{L}_{\text{OR}} = -\log \sigma(\log \text{OR}_\theta(y^w, y^l)) \tag{4}$$

and

$$\text{OR}_\theta(y^w, y^l) = \frac{\text{Odds}_\theta(y^w)}{\text{Odds}_\theta(y^l)}$$

is the odds ratio between preferred and non-preferred responses. The parameter α balances SFT and preference optimization.

4. MINT Framework

MINT (Multimodal Integrated Knowledge Transfer) is a framework designed to transfer domain-specific knowledge from multimodal models to unimodal LLMs through preference optimization. As shown in Figure 1, MINT consists of two main stages: (1) Upstream Preference Dataset Construction and (2) Downstream LLM Alignment. This framework enables unimodal models to benefit from multimodal expertise without requiring multimodal input during inference.

4.1. Upstream Pipeline

The upstream pipeline employs a multimodal model g_ϕ to convert raw multimodal data into a so-called *preference-learning dataset*. For each sample $i \in \{1, \ldots, N\}$ we observe $(x_i^{\text{text}}, x_i^{\text{image}}, y_i^*)$, where x_i^{text} and x_i^{image} denote the text and image modalities, respectively, and y_i^* is the ground-truth label. Feeding the multimodal pair into g_ϕ yields a predictive distribution $p_\phi\left(y | x_i^{\text{text}}, x_i^{\text{image}}\right)$.

Preferred & non-preferred label sets. From this distribution we extract the *top-k* labels $Y_i^w = \text{Top-}k\big(p_\phi(\cdot \mid x_i^{\text{text}}, x_i^{\text{image}})\big)$ as **preferred** responses and the *bottom-q* labels $Y_i^l = \text{Bottom-}q(\cdot)$ as **non-preferred** responses. If $y_i^* \notin Y_i^w$ we replace the lowest-ranked element of Y_i^w with y_i^* and shuffle it to the top-rank to ensure the ground truth is always preferred. We then form every ordered pair $(y^w, y^l) \in Y_i^w \times Y_i^l$ and attach it to the unimodal input $x_i^{\text{uni}} \equiv x_i^{\text{text}}$ to obtain triples $\left(x_i^{\text{uni}}, y^w, y^l\right)$, which constitute the preference dataset D_{pref} used in Alg. 1.

Acceptance-over-Rejection ratio. The hyper-parameter choice (k, q) controls the *Acceptance-over-Rejection* ratio $\text{AoR} = k/q$, balancing positive and negative feedback and thereby influencing the downstream ORPO learning dynamics.

Algorithm 1 MINT Framework

Require: Multimodal model g_ϕ, LLM f_θ, dataset $\{(x_i^{\text{text}}, x_i^{\text{image}}, y_i^*)\}_{i=1}^N$;
top-k, bottom-q, ORPO weight λ, learning rate η;
Phase 1 Upstream Pipeline: // *Construct preference dataset*
$D_{\text{pref}} \leftarrow \emptyset$;
for $i = 1$ to N **do**
  $p_\phi(y \mid x_i^{\text{text}}, x_i^{\text{image}}) \leftarrow g_\phi(x_i^{\text{text}}, x_i^{\text{image}})$;
  $Y_i^w \leftarrow$ top-k labels from $p_\phi(y \mid x_i^{\text{text}}, x_i^{\text{image}})$;
  $Y_i^l \leftarrow$ bottom-q labels from $p_\phi(y \mid x_i^{\text{text}}, x_i^{\text{image}})$;
  if $y_i^* \notin Y_i^w$ **then**
    Replace lowest-ranked label in Y_i^w with y_i^*;
  end
  for $y^w \in Y_i^w$ **do**
    for $y^l \in Y_i^l$ **do**
      Add $(x_i^{\text{uni}}, y^w, y^l)$ to D_{pref};
    end
  end
end
Phase 2 Downstream Pipeline: // *ORPO training on preference data*
while *not converged* **do**
  Sample batch $\mathcal{B} \subset D_{\text{pref}}$;
  $\mathcal{L}_{\text{SFT}} \leftarrow -\frac{1}{|\mathcal{B}|} \sum_{(x, y^w, y^l) \in \mathcal{B}} \log p_\theta(y^w \mid x)$;
  for $(x, y^w, y^l) \in \mathcal{B}$ **do**
    $p_w \leftarrow p_\theta(y^w \mid x)$; $p_l \leftarrow p_\theta(y^l \mid x)$;
    $\text{OR}_\theta(x, y^w, y^l) \leftarrow \frac{p_w/(1-p_w)}{p_l/(1-p_l)}$;
  end
  $\mathcal{L}_{\text{OR}} \leftarrow -\frac{1}{|\mathcal{B}|} \sum \log \sigma(\log \text{OR}_\theta(x, y^w, y^l))$;
  $\mathcal{L}_{\text{ORPO}} \leftarrow \mathcal{L}_{\text{SFT}} + \lambda \cdot \mathcal{L}_{\text{OR}}$;
  $\theta \leftarrow \theta - \eta \cdot \nabla_\theta \mathcal{L}_{\text{ORPO}}$;
end

4.2. Downstream Pipelines

To benchmark our method, we first consider *standard Supervised Fine-tuning* (SFT, top-right of Fig. 1). Given the preference dataset D_{pref}, SFT minimizes $\mathcal{L}_{\text{SFT}} = -\log p_\theta(y^w \mid x)$ on the preferred labels only. Although easy to implement, SFT teaches the LLM *what to predict* but ignores *what to avoid*, since non-preferred labels y^l never appear in the objective.

MINT with DPO. In the MINT-DPO variant (bottom-left of Fig. 1) we first obtain a *reference model* $f_{\theta_{\text{ref}}}$ by SFT. Subsequently, we optimize a *policy model* f_θ with the DPO loss introduced in Section 3 Equation (2) thereby contrasting preferred (y^w) and nonpreferred (y^l) labels while the reference model regularizes the update.

MINT with ORPO (default). Our default implementation, MINT-ORPO (bottom right of Fig. 1), eliminates the need

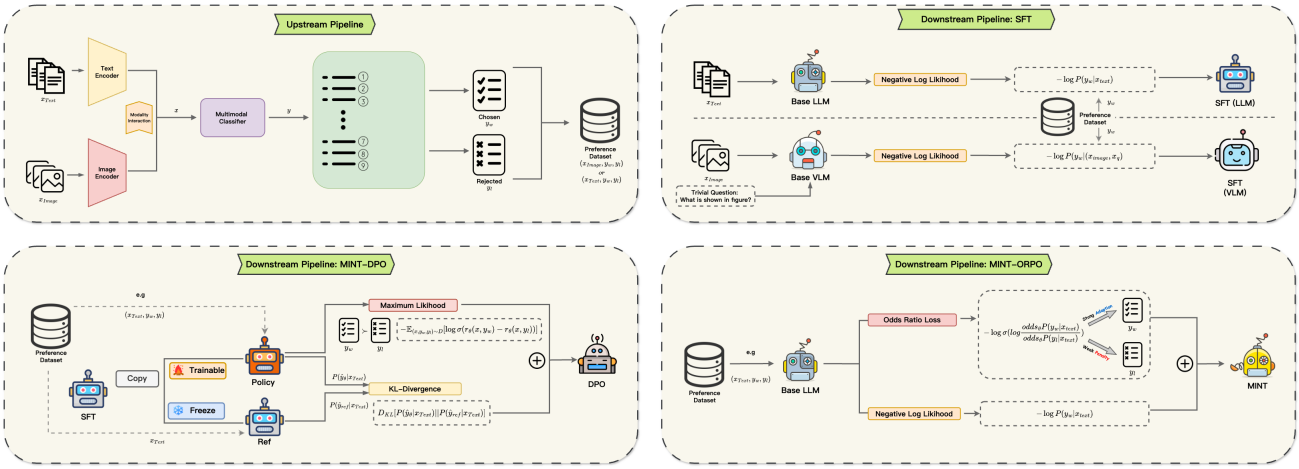


Figure 1. Overview of the MINT (Multimodal Integrated Knowledge Transfer) framework. (Top-Left) Upstream Pipeline: A multimodal model generates preferred (top-k) and non-preferred (bottom-q) responses from multimodal inputs, creating a preference dataset. (Top-Right) Baseline: Traditional Supervised Fine-Tuning (SFT) approach. (Bottom) MINT implementations: The preference dataset is used to align unimodal LLMs through either Direct Preference Optimization (DPO, Bottom-Left) or Odds Ratio Preference Optimization (ORPO, Bottom-Right), with ORPO serving as our default implementation.

for a separate reference model by jointly optimizing the single-stage objective introduced in Section 3 Equation (3), where the odds-ratio term (4) provides both positive and negative feedback and empirically produces stabler gradients when $|D_{\text{pref}}|$ is small.

Take-away. Both MINT-DPO and MINT-ORPO exploit the full preference data set $\left(x^{\text{uni}}, y^{w}, y^{l}\right) \in D_{\text{pref}}$, allowing the LLM to learn not only what is correct, but also which alternatives must be down-weighted, a capability absent from plain SFT. The complete training routine is summarized in Algorithm 1.

4.3. Biomedical Applications

We apply MINT to two challenging biomedical tasks:

For rare disease prediction, we use GestaltMML as the upstream multimodal model, with $k = q = 10$. The GestaltMML processes $(x_i^{\text{text}}, x_i^{\text{image}})$ pairs containing clinical phenotypes and facial images, while the downstream LLM makes predictions using only x_i^{text}.

For tissue type classification, we use PLIP as the upstream multimodal model, with $k = q = 5$. While the downstream model is technically a vision-language model (Llama 3.2-Vision-11B-Instruct), we effectively treat this as an image-only task. This is because the text input during inference is merely a generic prompt (e.g., "What is shown in this figure?") that provides no task-specific information. The model must rely entirely on the visual features from cell nucleus images to classify tissue types.

For both tasks, we convert the classification problem into a text generation task, enabling evaluation using standard metrics like Top-N accuracy. Table 2 and Table 3 shows examples of the prompts used for these tasks. Complete prompt specifications can be found in the supplementary material A.2.

4.4. Bias Mitigation Considerations.

To mitigate potential biases in preference dataset construction, we implement several safeguards: (1) we ensure balanced representation across different demographic groups in our training data; (2) we use multiple random seeds for all experiments to account for initialization-dependent biases; (3) we validate our approach on external datasets with different distributions. However, we acknowledge that systematic biases may still exist and recommend future work to develop more robust preference elicitation methods that account for uncertainty and demographic fairness.

5. Experiments

5.1. Experiment Setup

We evaluate MINT on two biomedical tasks: (1) Rare disease prediction from clinical text descriptions and (2) Tissue type classification from histological nucleus images.

Datasets. For rare disease prediction, we use the GestaltMatcher Database (GMDB) (Lesmann et al., 2023) with 6,522 training ad 386 in-distribution testing samples. For out-of-distribution external validation, we use 5,980 Phenopacket-derived clinical notes (Wu et al., 2025) split into overlapping and disjoint disease subsets compared to

GMDB disease types. For tissue type classification, we use the PanNuke dataset (Gamper et al.; 2020) with 5,436 training and 2,330 testing images across 19 tissue types.

Models. For rare disease prediction, we primarily use Llama 3.2-3B-Instruction as the downstream model and Gestalt-MML as the upstream model, which serves as our main experimental setting. For tissue type classification, we use Llama 3.2-Vision-11B-Instruct with PLIP as the upstream model. We also evaluate MINT on a wide range of models with varying sizes and architectures, including Llama 3.2 (1B, 3B), Gemma-2 (2B, 9B), and Llama 3.1 (8B, 70B, 405B). Due to computational constraints, only inference is performed on the largest models (70B, 405B). Full results across all model sizes and architectures are presented in Table 1 in the Appendix.

Implementation Details. We use LoRA fine-tuning (Hu et al., 2021) with rank 128 and scaling factor 64 for 3B/11B models, and rank 256/scaling factor 128 for 1B models, with dropout 0.05. We train all models with a learning rate of $5e^{-5}$. For SFT, we train for 5 epochs; for MINT with DPO, we first perform SFT for 5 epochs and then DPO for an additional 2 epochs; for MINT with ORPO (our default implementation), we train for 5 epochs. For preference datasets, we set $k = q = 10$ for rare disease prediction and $k = q = 5$ for tissue classification as default settings. All experiments are conducted on 4 NVIDIA A100 GPUs with 40GB memory each.

Baselines. We compare MINT against: (1) Base models without enhancement; (2) RAG (Lewis et al., 2020) (for rare disease only); (3) SFT (Gunel et al., 2020); and (4) MINT with DPO (Rafailov et al., 2024).

RAG Implementation. For the rare disease prediction task, we implement RAG using LlamaIndex (Liu, 2022) to construct our retrieval database. The database is built from two key biomedical resources: OMIM (Online Mendelian Inheritance in Man) (Zeeberg et al., 2005) and HPO (Human Phenotype Ontology) (Zhong et al., 2004). To improve retrieval precision, we preprocess these datasets by systematically mapping HPO and OMIM terms to their corresponding disease names, ensuring that phenotypic and disease-related terminology is accurately aligned.

To enhance the embedding representations for diseases and phenotypes, we adopt *NeuML/pubmedbert-base-embeddings* (Gu et al., 2021), a biomedical sentence transformer specifically trained on large-scale biomedical literature. This model captures the complex relationships and semantics of medical terms, allowing our system to understand the nuances of disease and phenotype descriptions more effectively. During inference, the RAG system retrieves relevant disease information based on the input clinical phenotypes and provides this context to the LLM for generating predictions. Note that RAG is only implemented for the rare disease prediction task, as the tissue type classification task relies primarily on visual features that are not well-represented in text-based knowledge bases.

Evaluation Metrics.

We assess model performance using four key metrics:

Hallucination-Free Accuracy (HFA). A response is considered non-hallucinatory if it contains exactly m labels in the required format and all labels match known entries in our reference database with similarity score ≥ 0.6, where $m = 10$ for rare disease prediction and $m = 5$ for tissue type classification.

Top-N Accuracy. After passing the HFA check, a response is correct for Top-N accuracy if the ground truth label appears among the N generated labels with similarity score ≥ 0.8. We use $N = 10$ for rare disease prediction and $N = 5$ for tissue type classification.

Top-1 Accuracy. After passing the HFA check, a response is correct for Top-1 accuracy if the first (highest-ranked) generated label matches the ground truth with similarity score ≥ 0.8.

Coverage-Avoidance Ratio (CAR). For response R_i, preferred labels Y_i^w, and non-preferred labels Y_i^l (from Algorithm 1), we define:

$$C_{k,i} = \frac{|Y_i^w \cap R_i|}{|R_i|}, \quad A_{q,i} = 1 - \frac{|R_i \cap Y_i^l|}{|R_i|}$$

$$\text{CAR}_{(k,q)} = \frac{1}{N} \sum_{i=1}^{N} \frac{2C_{k,i} \cdot A_{q,i}}{C_{k,i} + A_{q,i}}$$

where $C_{k,i}$ measures coverage of preferred labels, $A_{q,i}$ measures avoidance of non-preferred labels, and CAR is their harmonic mean.

SequenceMatcher. Python algorithm to compute string similarity ratios between 0 and 1 based on longest contiguous matching subsequences, handling variations in naming conventions (e.g., "Williams-Beuren Syndrome" vs. "Williams Beuren Syndrome").

6. Results

Table 1 presents our comprehensive evaluation results for both rare disease prediction and tissue type classification tasks across different enhancement techniques. We analyze these results in detail below.

6.1. Rare Disease Prediction

Main Results. On the main GMDB dataset, MINT consistently outperforms all baselines across all metrics. The base

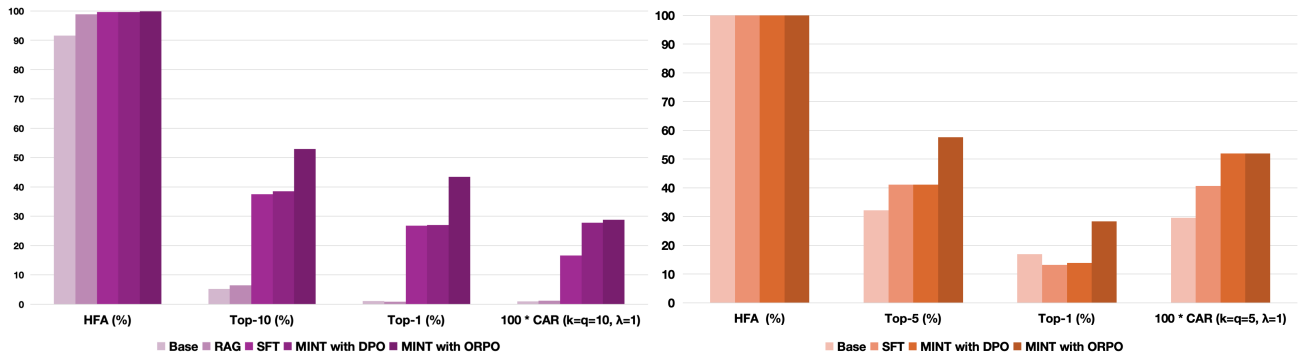


Figure 2. Performance comparison of different enhancement techniques. (Left) Rare disease prediction performance showing Top-10 accuracy, Top-1 accuracy, and CAR across all methods. (Right) Tissue type classification performance showing Top-5 accuracy, Top-1 accuracy, and CAR. MINT consistently outperforms all baselines across both tasks.

Llama 3.2-3B-Instruct model demonstrates limited ability to predict rare diseases, achieving only 5.19% Top-10 accuracy. RAG offers minimal improvement (6.52%), suggesting that simply retrieving relevant information without properly integrating it into the model's reasoning process is insufficient for this complex task.

SFT substantially improves performance (37.53% Top-10 accuracy), while DPO provides a slight additional gain (38.49% Top-10 accuracy). MINT significantly outperforms these approaches, achieving 52.99% Top-10 accuracy, representing a 37.7% relative improvement over DPO. The improvement in Top-1 accuracy is even more substantial (43.49% for MINT vs. 27.07% for DPO), indicating MINT's superior ability to identify the most likely diagnosis.

The CAR metric further highlights MINT's effectiveness (0.2877 vs. 0.2780 for DPO and 0.1665 for SFT), showing that MINT not only improves recall of correct diseases but also better avoids implausible ones. All enhancement techniques maintain high HFA values ($\geq$98.9%), with MINT achieving the highest at 99.91%.

External Validation. To assess generalization capabilities, we evaluate all methods on Phenopacket-derived clinical notes, split into overlapping diseases (those seen during training) and disjoint diseases (unseen during training). For overlapping diseases, MINT outperforms all baselines with 66.91% Top-10 accuracy and 47.56% Top-1 accuracy, showing even stronger relative improvements than on the main dataset.

For disjoint diseases (zero-shot scenario), all preference-based methods (SFT, DPO, MINT) show reduced performance, while RAG achieves the best results (24.17% Top-10 accuracy) compared to MINT (10.48%). This pattern reveals the complementary strengths of these approaches: MINT excels when it has encountered similar disease patterns during training, while RAG's retrieval capability provides an advantage for previously unseen conditions. This suggests that combining MINT with retrieval mechanisms could be particularly effective for comprehensive rare disease prediction systems.

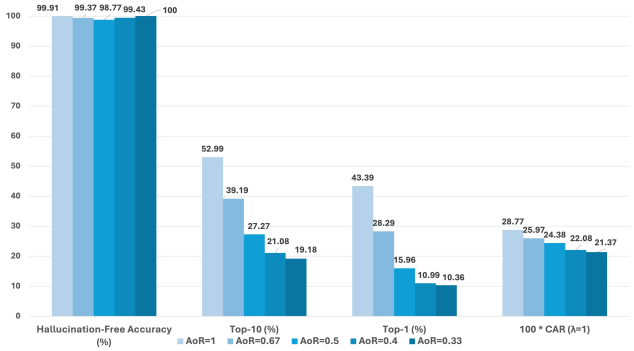


Figure 3. Effect of varying Acceptance-over-Rejection (AoR) ratios on MINT performance for rare disease prediction. A balanced ratio ($k/q = 10/10$) yields optimal performance across all metrics.

Impact of Acceptance-over-Rejection Ratio. Figure 3 shows how varying the balance between accepted and rejected samples affects MINT's performance. A balanced ratio ($k/q = 10/10$) yields the best performance (53.88% Top-10 accuracy), with accuracy decreasing as the ratio becomes more imbalanced (20.21% at 10/30). This suggests that a well-balanced preference dataset provides clearer and more stable gradient signals for learning disease ranking patterns.

6.2. Tissue Type Classification

In the tissue type classification task, MINT again demonstrates substantial improvements over all baselines. The base vision-language model achieves moderate performance (32.21% Top-5 accuracy), while SFT and DPO both improve to 41.46%. Interestingly, while SFT improves Top-5 accuracy, it slightly decreases Top-1 accuracy from 16.96% to 13.22%, suggesting that it helps identify plausible tissues

Table 1. Performance comparison across tasks and enhancement techniques. Results show Hallucination-Free Accuracy (HFA), Top-N Accuracy ($N = 10$ for rare disease, $N = 5$ for tissue type), Top-1 Accuracy, and Coverage-Avoidance Ratio (CAR). External validation results for rare disease prediction are shown for Phenopacket-derived clinical notes with overlapping and disjoint diseases. Bold values indicate best performance for each metric in each category.

Task	Method	HFA (%)↑	Top-N (%)↑	Top-1 (%)↑	CAR ↑
Rare Disease Prediction (GMDB Test Dataset)	Base Model	91.61	5.19	1.06	0.0522
	+RAG	98.90	6.52	0.91	0.0517
	+SFT	99.71	37.53	26.86	0.1665
	+MINT with DPO	99.63	38.49	27.07	0.2780
	+MINT with ORPO	**99.91**	**52.99**	**43.49**	**0.2877**
Rare Disease (Phenopacket-derived Overlapping Diseases)	Base Model	99.94	5.13	3.42	–
	+RAG	**100.00**	30.11	26.45	–
	+SFT	99.94	46.40	14.65	–
	+MINT with DPO	99.94	60.99	26.56	–
	+MINT with ORPO	99.89	**66.91**	**47.56**	–
Rare Disease (Phenopacket-derived Disjoint Diseases)	Base Model	99.99	20.00	11.29	–
	+RAG	**100.00**	**24.17**	**13.80**	–
	+SFT	100.00	9.16	8.46	–
	+MINT with DPO	99.99	9.48	7.71	–
	+MINT with ORPO	99.90	10.48	7.00	–
Tissue Type Classification (PanNuke Test Dataset)	Base Model	100.00	32.21	16.96	0.2965
	+SFT	100.00	41.46	13.22	0.4065
	+MINT with DPO	100.00	41.46	13.91	0.5196
	+MINT with ORPO	**100.00**	**57.58**	**28.41**	**0.5203**

but may struggle with precise ranking.

MINT significantly outperforms these approaches, achieving 57.58% Top-5 accuracy (a 38.9% relative improvement over SFT/DPO) and 28.41% Top-1 accuracy (a 104.2% relative improvement over the base model). The CAR metric shows similar patterns, with MINT achieving the highest value (0.5203), slightly outperforming DPO (0.5196) and substantially exceeding SFT (0.4065). All methods maintain perfect HFA values (100%), indicating that hallucination is not a significant issue for this task.

Discriminating Similar Tissues. Figure 4 presents a case study comparing the models' ability to distinguish between histologically similar bile duct and colon tissues. For bile duct samples, MINT consistently assigns rank 1 to the correct tissue while relegating the visually similar colon tissue to lower ranks (average rank 5.75). In contrast, SFT shows considerable confusion, assigning average ranks of 1.25 to bile duct (correct) and 2.00 to colon (incorrect).

Similarly, for colon samples, MINT assigns rank 1 to the correct tissue and pushes bile duct to much lower positions (average rank 4.00), while SFT shows significant confusion with average ranks of 1.25 for colon (correct) and 1.50 for bile duct (incorrect). This demonstrates MINT's superior ability to learn subtle discriminative features between visually similar tissues through its preference optimization approach.

6.3. Preservation of General Capabilities

A potential concern with domain-specific enhancement is whether it compromises a model's general capabilities. Figure 5 shows the performance of different approaches on general language and vision benchmarks. For rare disease prediction models, we evaluate on six language understanding benchmarks: MMLU, TruthfulQA, HellaSwag, Winogrande, ARC, and GSM8k. For tissue classification models, we use SEED-Bench, a comprehensive multimodal benchmark.

The results confirm that all enhancement techniques, including MINT, maintain performance comparable to the base models across all benchmarks. This indicates that MINT's domain-specific improvements do not come at the expense of general language or visual understanding capabilities, a crucial property for models that may be deployed in diverse real-world settings.

6.4. Theoretical Understanding

While our work provides strong empirical evidence for MINT's effectiveness, we acknowledge that the theoretical understanding of why preference optimization successfully transfers multimodal knowledge to unimodal models warrants further investigation. We hypothesize that preference optimization enables the model to learn not just the mapping from inputs to outputs, but the underlying decision boundaries that distinguish between correct and incorrect

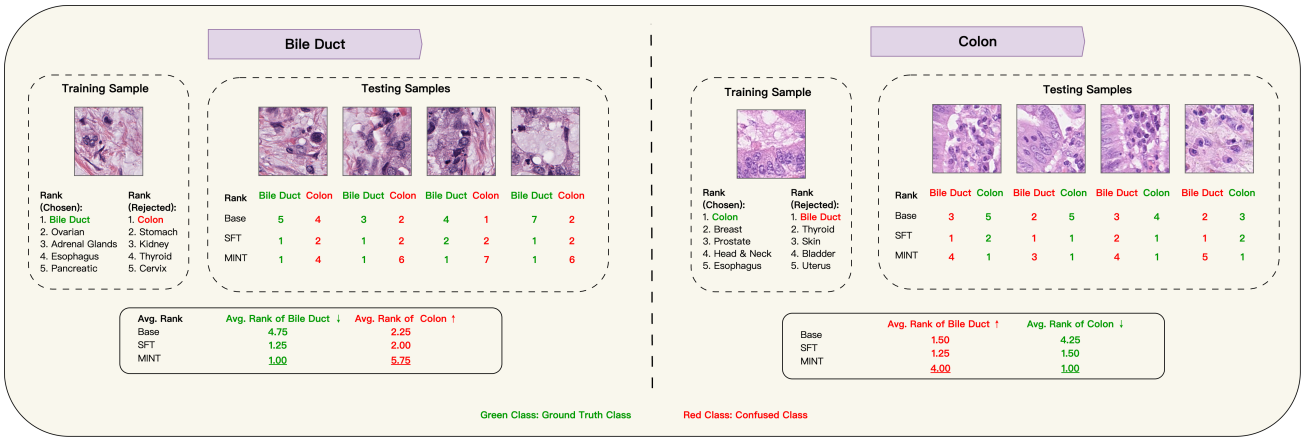


Figure 4. Comparative analysis of tissue type classification performance between Base model, SFT, and MINT for similar-looking bile duct and colon tissues. Green values represent rankings for ground truth tissue class, red values for visually similar confused class.

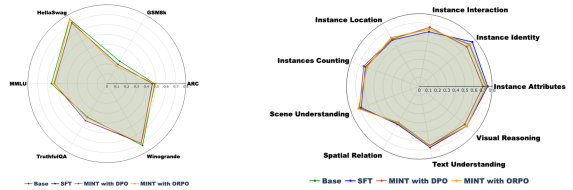


Figure 5. Preservation of general capabilities. Left: Performance on six language understanding benchmarks for rare disease prediction models. Right: Performance on SEED-Bench for tissue classification models.

predictions. When the upstream multimodal model generates preferences based on its multimodal understanding, these preferences encode the discriminative patterns learned from both text and image modalities. Through preference optimization, the downstream unimodal model internalizes these decision patterns, effectively inheriting the multimodal model's ability to distinguish between similar cases even when processing only a single modality.

The connection between preference learning and knowledge transfer can be understood through the lens of implicit knowledge distillation. Unlike traditional knowledge distillation that transfers soft probabilities, MINT transfers ranked preferences, which may preserve more nuanced decision-making patterns. Future theoretical work could formalize this connection and provide bounds on the knowledge transfer efficacy. The detailed analysis is included in A.7.

7. Conclusion

In this study, we introduced MINT (Multimodal Integrated Knowledge Transfer), a framework for aligning unimodal LLMs with domain-specific knowledge from multimodal biomedical sources through preference optimization. Our experiments on rare disease prediction and tissue type classification demonstrate MINT's effectiveness, improving Top-10 accuracy from 5.19% to 52.99% for rare disease prediction and Top-5 accuracy from 32.21% to 57.58% for tissue classification, outperforming both SFT and DPO approaches. We found that a balanced ratio of preferred and non-preferred examples provides optimal learning signals, and that MINT excels at distinguishing between similar cases (e.g., histologically similar tissues) - a critical capability for clinical applications. While MINT performs best on in-distribution samples, combining it with retrieval mechanisms could address zero-shot scenarios with previously unseen classes. MINT offers advantages over existing approaches: compared to SFT, it enables more nuanced discrimination between similar cases; compared to RAG, it internalizes domain knowledge during training, eliminating inference-time latency. Future work includes exploring hybrid approaches with retrieval mechanisms, extending to additional biomedical tasks, and investigating interpretability techniques to understand the patterns learned through preference optimization.

Impact Statement

This work aims to advance machine learning applications in healthcare by enabling more accurate rare disease prediction and tissue type classification from limited modalities. While MINT offers potential benefits for clinical diagnostic workflows, particularly in resource-constrained settings, we acknowledge that any AI-based diagnostic system should serve as a supportive tool for healthcare professionals rather than a standalone solution. Future deployment should address potential demographic biases in training data and maintain transparency about model capabilities and limitations

to ensure equitable performance across diverse populations.

References

Afantenos, S., Karkaletsis, V., and Stamatopoulos, P. Summarization from medical documents: a survey. *Artificial intelligence in medicine*, 33(2):157–177, 2005. ISSN 0933-3657.

Al-Moslmi, T., Ocaña, M. G., Opdahl, A. L., and Veres, C. Named entity extraction for knowledge graphs: A literature overview. *IEEE Access*, 8:32862–32881, 2020. ISSN 2169-3536.

Berglund, L., Tong, M., Kaufmann, M., Balesni, M., Stickland, A. C., Korbak, T., and Evans, O. The reversal curse: Llms trained on" a is b" fail to learn" b is a". *arXiv preprint arXiv:2309.12288*, 2023.

Chen, T., Liu, S., Chang, S., Cheng, Y., Amini, L., and Wang, Z. Adversarial robustness: From self-supervised pre-training to fine-tuning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 699–708.

Daiber, J., Jakob, M., Hokamp, C., and Mendes, P. N. Improving efficiency and accuracy in multilingual entity extraction. In *Proceedings of the 9th international conference on semantic systems*, pp. 121–124.

do Olmo, J., Logrono, J., Mascias, C., Martinez, M., and Isla, J. Assessing dxgpt: Diagnosing rare diseases with various large language models. *medRxiv*, pp. 2024.05. 08.24307062, 2024.

Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Yang, A., and Fan, A. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

Engstrom, L., Ilyas, A., Santurkar, S., Tsipras, D., Janoos, F., Rudolph, L., and Madry, A. Implementation matters in deep rl: A case study on ppo and trpo. In *International conference on learning representations*.

Engstrom, L., Ilyas, A., Santurkar, S., Tsipras, D., Janoos, F., Rudolph, L., and Madry, A. Implementation matters in deep policy gradients: A case study on ppo and trpo. *arXiv preprint arXiv:2005.12729*, 2020.

Etzioni, O., Cafarella, M., Downey, D., Popescu, A.-M., Shaked, T., Soderland, S., Weld, D. S., and Yates, A. Unsupervised named-entity extraction from the web: An experimental study. *Artificial intelligence*, 165(1):91–134, 2005. ISSN 0004-3702.

Gamper, J., Alemi Koohbanani, N., Benet, K., Khuram, A., and Rajpoot, N. Pannuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification. In *Digital Pathology: 15th European Congress, ECDP 2019, Warwick, UK, April 10–13, 2019, Proceedings 15*, pp. 11–19. Springer. ISBN 3030239365.

Gamper, J., Koohbanani, N. A., Benes, K., Graham, S., Jahanifar, M., Khurram, S. A., Azam, A., Hewitt, K., and Rajpoot, N. Pannuke dataset extension, insights and baselines. *arXiv preprint arXiv:2003.10778*, 2020.

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., and Wang, H. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*, 2023.

Gu, Y., Tinn, R., Cheng, H., Lucas, M., Usuyama, N., Liu, X., Naumann, T., Gao, J., and Poon, H. Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare (HEALTH)*, 3(1):1–23, 2021. ISSN 2691-1957.

Gunel, B., Du, J., Conneau, A., and Stoyanov, V. Supervised contrastive learning for pre-trained language model fine-tuning. *arXiv preprint arXiv:2011.01403*, 2020.

Hinton, G., Vinyals, O., and Dean, J. Distilling the knowledge in a neural network, 2015. URL https://arxiv.org/abs/1503.02531.

Hong, J., Lee, N., and Thorne, J. Orpo: Monolithic preference optimization without reference model. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 11170–11189.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., and Chen, W. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.

Huang, Z., Bianchi, F., Yuksekgonul, M., Montine, T. J., and Zou, J. A visual–language foundation model for pathology image analysis using medical twitter. *Nature medicine*, 29(9):2307–2316, 2023. ISSN 1078-8956.

Jiang, Z., Xu, F. F., Gao, L., Sun, Z., Liu, Q., Dwivedi-Yu, J., Yang, Y., Callan, J., and Neubig, G. Active retrieval augmented generation. *arXiv preprint arXiv:2305.06983*, 2023.

Kim, J., Wang, K., Weng, C., and Liu, C. Assessing the utility of large language models for phenotype-driven gene prioritization in the diagnosis of rare genetic disease. *The American Journal of Human Genetics*, 111(10): 2190–2202, 2024. ISSN 0002-9297. doi: 10.1016/j.ajhg. 2024.08.010. URL https://doi.org/10.1016/j.ajhg.2024.08.010.

Lesmann, H., Lyon, G. J., Caro, P., Abdelrazek, I. M., Moosa, S., Pantel, J. T., Ten Hagen, M., Rosnev, S., Kamphans, T., and Meiswinkel, W. Gestaltmatcher database-a fair database for medical imaging data of rare disorders. *MedRxiv*, 2023.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., and Rocktäschel, T. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, 33:9459–9474, 2020.

Liu, J. Llamaindex, 2022. URL https://github.com/jerryjliu/llama_index.

Mamykina, L., Vawdrey, D. K., Stetson, P. D., Zheng, K., and Hripcsak, G. Clinical documentation: composition or synthesis? *Journal of the American Medical Informatics Association*, 19(6):1025–1031, 2012. ISSN 1527-974X.

McKinzie, B., Gan, Z., Fauconnier, J.-P., Dodge, S., Zhang, B., Dufter, P., Shah, D., Du, X., Peng, F., and Belyi, A. Mm1: methods, analysis and insights from multimodal llm pre-training. In *European Conference on Computer Vision*, pp. 304–323. Springer.

Meta. Llama 3.2 model card, 2024a. URL models/llama3_2/MODEL_CARD.md.

Meta. Llama 3.3 model card, 2024b. URL models/llama3_3/MODEL_CARD.md.

Qiu, J., Lam, K., Li, G., Acharya, A., Wong, T. Y., Darzi, A., Yuan, W., and Topol, E. J. Llm-based agentic systems in medicine and healthcare. *Nature Machine Intelligence*, 6(12):1418–1420, 2024. ISSN 2522-5839.

Rafailov, R., Sharma, A., Mitchell, E., Manning, C. D., Ermon, S., and Finn, C. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36, 2024.

Tripathi, S., Sukumaran, R., and Cook, T. S. Efficient healthcare with large language models: optimizing clinical workflow and enhancing patient care. *Journal of the American Medical Informatics Association*, 31(6): 1436–1440, 2024. ISSN 1067-5027.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, , and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., and Brew, J. Transformers: State-of-the-art natural language processing. In *Conference on Empirical Methods in Natural Language Processing*, 2020.

Wu, D., Yang, J., Liu, C., Hsieh, T.-C., Marchi, E., Blair, J., Krawitz, P., Weng, C., Chung, W., and Lyon, G. J. Gestaltmml: Enhancing rare genetic disease diagnosis through multimodal machine learning combining facial images and clinical texts. *ArXiv*, 2024a.

Wu, D., Yang, J., and Wang, K. Exploring the reversal curse and other deductive logical reasoning in bert and gpt-based large language models. *Patterns*, 5(9), 2024b. ISSN 2666-3899.

Wu, D., Wang, Z., Nguyen, Q., and Wang, K. Integrating chain-of-thought and retrieval augmented generation enhances rare disease diagnosis from clinical notes. *arXiv preprint arXiv:2503.12286*, 2025.

Yu, C., Velu, A., Vinitsky, E., Gao, J., Wang, Y., Bayen, A., and Wu, Y. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in neural information processing systems*, 35:24611–24624, 2022.

Zeeberg, B. R., Qin, H., Narasimhan, S., Sunshine, M., Cao, H., Kane, D. W., Reimers, M., Stephens, R. M., Bryant, D., Burt, S. K., Elnekave, E., Hari, D. M., Wynn, T. A., Cunningham-Rundles, C., Stewart, D. M., Nelson, D., and Weinstein, J. N. High-throughput gominer, an 'industrial-strength' integrative gene ontology tool for interpretation of multiple-microarray experiments, with application to studies of common variable immune deficiency (cvid). *BMC Bioinformatics*, 6:168, 2005. URL http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15998470.

Zhong, S., Storch, K. F., Lipan, O., Kao, M. C., Weitz, C. J., and Wong, W. H. Gosurfer: a graphical interactive tool for comparative analysis of large gene sets in gene ontology space. *Appl Bioinformatics*, 3(4):261–4, 2004. URL http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15702958.

Zhu, H., Huang, B., Zhang, S., Jordan, M., Jiao, J., Tian, Y., and Russell, S. J. Towards a theoretical understanding of the'reversal curse'via training dynamics. *Advances in Neural Information Processing Systems*, 37:90473–90513, 2024.

A. Additional Experimental Details

A.1. Full Implementation Details

All experiments were implemented using PyTorch and the Hugging Face Transformers library. For LoRA fine-tuning, we used the PEFT library with the following configuration:

- LoRA rank: 128 for 3B/11B models, 256 for 1B models
- LoRA alpha: 64 for 3B/11B models, 128 for 1B models
- LoRA dropout: 0.05
- LoRA target modules: query, key, value, and output projection matrices
- Training precision: bfloat16

For training, we used the AdamW optimizer with a learning rate of $5e^{-5}$ and a cosine learning rate schedule with 500 warm-up steps. We applied gradient clipping with a maximum norm of 1.0 to prevent gradient explosion. For both SFT and MINT with ORPO, we trained for 5 epochs, while for MINT with DPO, we first performed SFT for 5 epochs and then DPO for an additional 2 epochs.

For ORPO training, we set the weighting factor λ (controlling the balance between SFT loss and odds ratio loss) to 1.0. For DPO, we set β (controlling KL divergence from the reference model) to 0.1.

We created preference datasets by extracting predictions from the upstream multimodal models (GestaltMML and PLIP). For rare disease prediction, we used Nvidia TensorRT for inference acceleration of the GestaltMML model. For tissue type classification, we integrated the official PLIP implementation for generating predictions.

A.2. Prompt Templates and Examples

Tables 2 and 3 provide detailed information about the prompts and examples used in our experiments.

Table 2. Prompt templates for rare disease prediction and tissue type classification tasks.

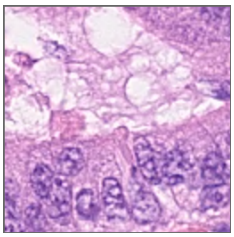
Component	Description
Rare Disease Diagnosis System Prompt	You are a genetic counselor. Your task is to identify potential rare diseases based on given phenotypes. Follow the output format precisely.
Rare Disease Diagnosis User Prompt	A [AGE] years old [GENDER] patient of [ETHNICITY] descent exhibits the following phenotypes: [PHENOTYPES]. What are the potential diagnoses? Based on this information, provide a numbered list of EXACTLY 10 potential rare diseases.
Tissue Type Classification System Prompt	You are a professional pathologist analyzing nucleus images. For each image, identify the most likely tissue types based on the nuclear morphology. Follow the output format precisely.
Tissue Type Classification User Prompt	Based on the morphological features observed in the provided nucleus image, which tissue types could this nucleus originate from? List exactly 5 tissue types in decreasing order of likelihood.
Preference Learning Format	[SYSTEM PROMPT] [USER PROMPT] Chosen Response: [CHOSEN RESPONSE] Rejected Response: [REJECTED RESPONSE]

These tables illustrate how our preference datasets were constructed from the outputs of the upstream multimodal models. For each task, we show the complete process from input data through multimodal model processing to the generation of preferred (chosen) and non-preferred (rejected) responses. The chosen responses contain the top predictions from the upstream model, while the rejected responses contain the bottom predictions, enabling effective preference learning.

A.3. Results Across Model Sizes

Table 4 presents detailed results across different model sizes and architectures.

Table 3. Examples showing the complete pipeline from multimodal model outputs to preference dataset construction.

Rare Disease Prediction	
Component	**Description**
Input	*A 3 year(s) old male patient with long palpebral fissure, growth delay, decreased body weight, eversion of lateral third of lower eyelids, feeding difficulties.*
Upstream Model Input	Text: Same as above. Image: Facial photograph
Upstream Model Output (Top-10)	Kabuki syndrome (0.28), Rubinstein-Taybi syndrome (0.19), Williams-Beuren syndrome (0.15), DiGeorge syndrome (0.12), Cornelia de Lange syndrome (0.09), etc.
Upstream Model Output (Bottom-10)	Bardet-Biedl syndrome (0.0004), Wolfram syndrome (0.0003), Marfan syndrome (0.0003), Waardenburg syndrome (0.0002), etc.
Chosen Response	*POTENTIAL_DISEASES: 1. Kabuki syndrome 2. Rubinstein-Taybi syndrome . . . 10. Cohen syndrome*
Rejected Response	*POTENTIAL_DISEASES: 1. Bardet-Biedl syndrome 2. Wolfram syndrome . . . 10. Ellis-van Creveld syndrome*
Tissue Type Classification	
Component	**Description**
Input	 *Nucleus histology image of colon tissue*
Upstream Model Input	Image: Same as above / Text: " What tissue type does this nucleus belong to?"
Upstream Model Output (Top-5)	Colon (0.35), Stomach (0.18), Breast (0.12), Prostate (0.08), Esophagus (0.06)
Upstream Model Output (Bottom-5)	Skin (0.004), Bile Duct (0.003), Thyroid (0.002), Liver (0.001), Adrenal Gland (0.001)
Chosen Response	*MOST_LIKELY_TISSUE: 1. Colon 2. Stomach 3. Breast 4. Prostate 5. Esophagus*
Rejected Response	*MOST_LIKELY_TISSUE: 1. Skin 2. Bile Duct 3. Thyroid 4. Liver 5. Adrenal Gland*

A.4. Effect of AoR Ratio

Table 5 shows detailed results for different Acceptance-over-Rejection (AoR) ratios on the rare disease prediction task.

A.5. General Capability Evaluation Details

To evaluate preservation of general capabilities, we used the following benchmarks:

Table 4. Performance of MINT across different model sizes and architectures on the rare disease prediction task. Standard deviations are computed across three random seeds.

Model	HFA (%)	Top-10 (%)	Top-1 (%)	CAR
Llama 3.2-1B-Instruct	95.04±1.15	4.24±0.23	0.65±0.18	0.0292±0.0016
+ MINT	**99.48±0.37**	**34.28±1.21**	**23.91±3.41**	**0.2120±0.0018**
Llama 3.2-3B-Instruct	91.61±0.59	5.19±0.01	1.06±0.02	0.0522±0.0025
+ MINT	**99.91±0.15**	**52.99±1.69**	**43.49±0.57**	**0.2877±0.0035**
Llama 3.1-8B-Instruct	95.07±1.45	6.63±0.21	1.67±0.17	0.0797±0.0005
+ MINT	**97.28±2.02**	**25.91±3.29**	**16.97±0.92**	**0.2326±0.0009**
Gemma-2-2B-Instruct	100.00±0.00	6.99±0.00	1.79±0.03	0.1065±0.0063
+ MINT	**99.04±1.36**	**16.32±0.00**	**8.04±0.01**	**0.1743±0.0051**
Gemma-2-9B-Instruct	99.56±0.63	12.37±0.82	5.03±0.16	0.1368±0.0047
+ MINT	**99.61±0.55**	**33.80±1.25**	**23.40±0.85**	**0.2465±0.0122**
Llama 3.1-70B-Instruct*	99.74±0.16	21.24±0.27	11.92±0.08	0.1122±0.0031
Llama 3.1-405B-Instruct*	100.00±0.00	26.17±0.22	11.66±0.08	0.1221±0.0019

*Due to computational constraints, only inference was performed on these models.

Table 5. Effect of varying Acceptance-over-Rejection (AoR) ratios on MINT performance for rare disease prediction using Llama 3.2-3B-Instruct. The k value is fixed at 10, while q varies from 10 to 30.

AoR Ratio (k/q)	HFA (%)	Top-10 (%)	Top-1 (%)	CAR
10/10 (1.0)	**100.00**	**53.88**	**44.04**	**1.63**
10/15 (0.67)	99.48	38.13	28.29	1.47
10/20 (0.5)	98.96	26.06	15.96	1.38
10/25 (0.4)	99.97	21.07	10.99	1.25
10/30 (0.33)	100.00	20.21	10.36	1.21

Language Benchmarks. For rare disease prediction models, we evaluated on six standard benchmarks:

- MMLU: 57 subjects across STEM, humanities, social sciences, with 14,042 multiple-choice questions
- TruthfulQA: 817 questions designed to test tendency to reproduce false claims
- HellaSwag: 10,042 commonsense inference questions with 4 choices each
- Winogrande: 1,267 commonsense reasoning questions based on Winograd schemas
- ARC: 7,787 grade-school science questions with challenging reasoning
- GSM8k: 1,319 grade-school math problems requiring multi-step solutions

We used 5-shot evaluation for all language benchmarks to provide a fair comparison across different models.

Vision-Language Benchmark. For tissue classification models, we used SEED-Bench which contains 19,000 multiple-choice questions across 12 evaluation dimensions. We report scores on the visual perception dimensions most relevant to the tissue classification task.

A.6. Dataset Details

GestaltMatcher Database (GMDB). The GMDB dataset contains facial images and clinical phenotypes for patients with genetic disorders. For our experiments, we used version 1.0.9, which includes 6,908 samples across ∼520 disease classes. We split the data into 6,522 training samples and 386 testing samples, ensuring no patient appears in both sets. The text input for each sample consists of demographic information (age, gender, ethnicity) and a list of HPO terms representing observed phenotypes.

Phenopacket-derived Clinical Notes. This external validation dataset consists of 5,980 synthetically generated clinical notes. Each note follows the structure of real clinical documentation and describes a patient case with observed phenotypes and the correct diagnosis. We divided this dataset into two subsets: 1,638 samples with 72 diseases that overlap with our training set, and 4,342 samples with 456 diseases that do not appear in our training set.

PanNuke Database. This dataset contains 481 visual fields of H&E-stained histology slides across 19 tissue types, with 205,343 labeled nuclei. We used the official train/test split, resulting in 5,436 training images and 2,330 testing images. Each image shows a single nucleus with surrounding context, and the task is to identify the tissue of origin from the 19 possible tissue types.

A.7. Theoretical Understanding

While our work provides strong empirical evidence for MINT's effectiveness, we provide theoretical insights into why preference optimization successfully transfers multimodal knowledge to unimodal models.

A.7.1. Knowledge Transfer Through Preference Learning

The effectiveness of MINT can be understood through three theoretical perspectives:

Implicit Multimodal Knowledge Distillation Traditional knowledge distillation transfers soft probabilities from teacher to student models (Hinton et al., 2015). MINT extends this concept by transferring **ranked preferences** that encode decision-making patterns learned from multimodal data. When the upstream multimodal model generates preference rankings based on joint text-image understanding, these rankings encapsulate the discriminative patterns that distinguish between correct and incorrect responses across modalities.

Formally, let $g_\phi(\mathbf{x}^{\text{text}}, \mathbf{x}^{\text{image}})$ be the multimodal teacher model that produces a predictive distribution over labels $\mathcal{Y}$. The preference ranking $\pi_\phi : \mathcal{Y} \to \mathbb{R}$ derived from this distribution contains implicit knowledge about the joint multimodal decision process:

$$\pi_\phi(y) = \log p_\phi(y|\mathbf{x}^{\text{text}}, \mathbf{x}^{\text{image}}) \tag{5}$$

Through preference optimization, the unimodal student model f_θ learns to approximate this ranking using only text input:

$$\pi_\theta(\mathbf{x}^{\text{text}}) \approx \pi_\phi(\mathbf{x}^{\text{text}}, \mathbf{x}^{\text{image}}) \tag{6}$$

This enables the student to make decisions that reflect multimodal understanding using only single-modality inputs.

Decision Boundary Learning Unlike supervised fine-tuning that only learns positive examples, MINT explicitly learns decision boundaries between correct and incorrect responses. The odds ratio in ORPO:

$$\text{OR}_\theta(y^w, y^l) = \frac{\text{Odds}_\theta(y^w)}{\text{Odds}_\theta(y^l)} = \frac{p_\theta(y^w)/(1 - p_\theta(y^w))}{p_\theta(y^l)/(1 - p_\theta(y^l))} \tag{7}$$

encourages the model to not only increase probability for preferred responses but simultaneously decrease probability for non-preferred ones. This contrastive learning enables more nuanced discrimination between similar cases—a critical capability in biomedical applications where subtle differences distinguish between similar diseases or tissue types.

The ORPO loss can be decomposed as:

$$\mathcal{L}_{\text{ORPO}} = \mathcal{L}_{\text{SFT}} + \alpha \cdot \mathcal{L}_{\text{OR}} \tag{8}$$

$$= -\log p_\theta(y^w|\mathbf{x}) - \alpha \log \sigma(\log \text{OR}_\theta(y^w, y^l)) \tag{9}$$

where the first term ensures the model learns to generate preferred responses, and the second term enforces relative preference ordering.

Multimodal Pattern Compression The preference dataset construction process acts as a compression mechanism that distills multimodal expertise into preference rankings. When the multimodal model processes $(\mathbf{x}^{\text{text}}, \mathbf{x}^{\text{image}})$ pairs, it learns complex feature interactions across modalities. The resulting preference rankings serve as a compressed representation of these interactions.

Let $\mathbf{h}^{\text{text}} = \text{Encoder}_{\text{text}}(\mathbf{x}^{\text{text}})$ and $\mathbf{h}^{\text{image}} = \text{Encoder}_{\text{image}}(\mathbf{x}^{\text{image}})$ be the encoded representations of text and image modalities. The multimodal model learns a joint representation:

$$\mathbf{h}^{\text{joint}} = f_{\text{fusion}}(\mathbf{h}^{\text{text}}, \mathbf{h}^{\text{image}}) \tag{10}$$

The preference ranking generated from this joint representation:

$$\pi_\phi = \text{Rank}(p_\phi(y|\mathbf{h}^{\text{joint}})) \tag{11}$$

contains information about both modalities. Through preference optimization, the unimodal model learns to approximate this ranking using only $\mathbf{h}^{\text{text}}$:

$$\pi_\theta(\mathbf{h}^{\text{text}}) \approx \pi_\phi(\mathbf{h}^{\text{joint}}) \tag{12}$$

A.7.2. Information-Theoretic Perspective

From an information-theoretic standpoint, MINT can be viewed as a method for preserving mutual information between multimodal inputs and outputs while reducing the input dimensionality.

Let $I(\mathbf{X}^{\text{text}}, \mathbf{X}^{\text{image}}; Y)$ denote the mutual information between multimodal inputs and the target labels. Traditional approaches that discard one modality suffer from information loss:

$$I(\mathbf{X}^{\text{text}}; Y) \leq I(\mathbf{X}^{\text{text}}, \mathbf{X}^{\text{image}}; Y) \tag{13}$$

MINT attempts to recover some of this lost information through the preference learning process. Specifically, the preference rankings π derived from multimodal models contain additional information:

$$I(\mathbf{X}^{\text{text}}, \Pi; Y) > I(\mathbf{X}^{\text{text}}; Y) \tag{14}$$

The effectiveness of knowledge transfer depends on how much of the multimodal mutual information can be captured in the preference rankings:

$$\eta = \frac{I(\mathbf{X}^{\text{text}}, \Pi; Y) - I(\mathbf{X}^{\text{text}}; Y)}{I(\mathbf{X}^{\text{text}}, \mathbf{X}^{\text{image}}; Y) - I(\mathbf{X}^{\text{text}}; Y)} \tag{15}$$

where $\eta \in [0, 1]$ represents the knowledge transfer efficiency.

A.7.3. Convergence Properties

The ORPO objective combines supervised learning and preference optimization, creating a multi-objective optimization problem. We can analyze the convergence properties of this combined objective.

Let θ^*_{SFT} be the optimal parameters for the SFT objective and θ^*_{OR} be the optimal parameters for the odds ratio objective. The ORPO objective seeks to find:

$$\theta^* = \arg\min_\theta \left[\mathcal{L}_{\text{SFT}}(\theta) + \alpha \mathcal{L}_{\text{OR}}(\theta)\right] \tag{16}$$

Under certain regularity conditions (convexity of individual losses, Lipschitz continuity), the combined objective converges to a point that balances both objectives. The parameter α controls the trade-off between learning preferred responses and learning preference rankings.

When $\alpha = 0$, we recover standard SFT: $\theta^* \to \theta^*_{\text{SFT}}$. When $\alpha \to \infty$, the optimization focuses primarily on preference ranking: $\theta^* \to \theta^*_{\text{OR}}$.

The optimal value of α depends on the quality of the preference dataset and the alignment between the SFT and preference objectives.

A.7.4. Theoretical Limitations and Future Directions

While this analysis provides insights into MINT's mechanisms, several theoretical questions remain:

Transfer Efficiency Bounds We lack formal bounds on η, the knowledge transfer efficiency. Establishing such bounds would require characterizing the relationship between text-image feature alignment and transferable information.

Preference Quality Measures No formal metrics exist to evaluate the quality of preference datasets for knowledge transfer. Developing such metrics would enable principled preference dataset construction.

Optimization Landscape The non-convex nature of neural network optimization means that the combined ORPO objective may have multiple local minima. Understanding the optimization landscape could inform better training strategies.

Generalization Bounds Traditional generalization bounds for supervised learning may not apply directly to preference-based learning. Developing preference-specific generalization bounds would strengthen theoretical foundations.

Future theoretical work should focus on: (1) establishing formal knowledge transfer guarantees, (2) characterizing the optimization landscape of combined objectives, and (3) developing information-theoretic frameworks for multimodal knowledge compression.