RADS: Reinforcement Learning-Based Sample Selection Improves Transfer Learning in Low-resource and Imbalanced Clinical Settings

Wei Han[1], David Martinez[1], Anna Khanina[3,4,5], Lawrence Cavedon[1], Karin Verspoor[1,2,3]*

[1]School of Computing Technologies, RMIT University
[2]School of Computing and Information Systems, The University of Melbourne
[3]National Centre for Infections in Cancer, Melbourne
[4]Department of Infectious Disease, Peter MacCallum Cancer Centre
[5]Sir Peter MacCallum Department of Oncology, The University of Melbourne

Abstract

A common strategy in transfer learning is few shot fine-tuning, but its success is highly dependent on the quality of samples selected as training examples. Active learning methods such as uncertainty sampling and diversity sampling can select useful samples. However, under extremely low-resource and class-imbalanced conditions, they often favor outliers rather than truly informative samples, resulting in degraded performance. In this paper, we introduce **RADS** (**R**einforcement **A**daptive **D**omain **S**ampling), a robust sample selection strategy using reinforcement learning (RL) to identify the most informative samples. Experimental evaluations on several real world clinical datasets show our sample selection strategy enhances model transferability while maintaining robust performance under extreme class imbalance compared to traditional methods. Our code is open-sourced on GitHub[1].

1 Introduction

Maximizing the utility of limited data is a crucial focus of Natural Language Processing (NLP) research in domains such as clinical texts where acquiring large amounts of gold standard data may be difficult due to data restrictions and the relative rarity of many disease conditions. The high cost of annotation in such highly specialized domains further limits availability of labeled data. Yet, the effectiveness of NLP techniques in healthcare heavily relies on the quality of annotated datasets, particularly because clinical data contains specialized symbols, abbreviations, and medical jargon (Touvron et al., 2023; Liu et al., 2024a).

Transfer Learning (TL) (Tan et al., 2018), in which knowledge learned from a task is reused to boost performance on a different but related (target) task, has shown effectiveness across various machine learning applications (Weiss et al., 2016) and opens new avenues for addressing low-resource scenarios. Previous works have attempted to leverage pretrained embeddings (Maimaiti et al., 2021) and few shot examples (Alyafeai et al., 2020) to facilitate transfer learning in NLP. However, when the target task offers very few labeled instances, these approaches may generate unreliable outputs. This is an especially acute problem in healthcare, where reliability is paramount.

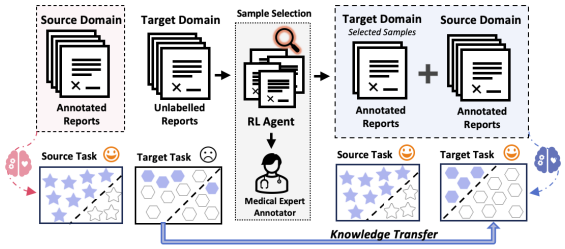

Figure 1: RL-based active sampling for transfer learning from source domain to target domain. Domain shift reduces zero-shot generalization from the source-trained model to the target domain. Our sample selection strategy uses RL to identify key samples from the target domain. By jointly fine-tuning on the selected target samples with the source data, the model achieves good performance on both domains.

Class imbalance (Johnson and Khoshgoftaar, 2019) is another challenge for low-resource settings. In clinical datasets, there is often a scarcity of positive cases due to the low prevalence of many conditions, making such instances both highly valuable and limited in number. At the same time, differences in data collection protocols can lead some disease datasets to contain a very high proportion of positive samples. These extreme disparities in class distribution further hinder the transferability of NLP models across clinical datasets.

Clinical documentation is heterogeneous, reflecting diverse investigations including CT and PET scans, or cytology and histopathology analysis. Although disease detection cues appear across these document types, their content structures, terminol-

*Corresponding author: karin.verspoor@rmit.edu.au
[1]https://github.com/Wei-0808/RADS

Findings of the Association for Computational Linguistics: ACL 2026, pages 12501–12516

ogy, and linguistic expressions can vary greatly. CT and PET scan reports primarily emphasize imaging-based findings (Townsend et al., 2004), whereas cytology and histopathology reports focus on cellular and tissue-level observations (Jensen, 2021).

Previous works have shown the efficacy of NLP techniques for disease detection from clinical reports. However, models fine-tuned on one report type show clear performance degradation when applied to another (Han et al., 2025). While disease-related signals overlap to some extent across different document types, existing disease detection models still fall short of human performance in transferring knowledge between them. As the preparation of gold standard annotated datasets for training is time-consuming, it is therefore important to explore effective knowledge transfer strategies from existing datasets to new but similar tasks. This not only improves annotation efficiency but also enhances the models' adaptability in dealing with variations in task settings.

In this work, we propose RADS (Reinforcement Adaptive Domain Sampling), a robust strategy for knowledge transfer between related but distinct sources. Following the active learning paradigm (Fu et al., 2013), we enhance transfer learning by identifying and selecting the most relevant samples for few-shot fine-tuning as shown in Figure 1. First, we employ an RL-based agent to identify the most informative samples within the target dataset. These selected samples are then annotated by medical experts and incorporated into the fine-tuning process. By jointly fine-tuning the model on the source dataset and the newly annotated target samples, the model is able to preserve strong performance on the source domain while achieving improved generalization to the target domain. We evaluated this approach across multiple real world clinical datasets. Experimental results show that our method improves both the adaptability and performance of disease detection between different sources. In the context of transfer learning, this technique offers a promising way to both reduce annotation effort and enhance model robustness in low-resource and class-imbalanced settings.

Our contributions are summarized as follows:

- This work addresses the challenges posed by low-resource and class-imbalance scenarios in disease detection across heterogeneous clinical report types from real-world clinical data sources.
- We propose RADS, a robust RL-based sample selection strategy tailored to scenarios with both data scarcity and class imbalance.
- Extensive experiments on several clinical datasets confirm that our transfer learning approach is more effective between similar but different sources, even under low-resource and class-imbalanced conditions.

2 Related Work

With high-quality annotated datasets, NLP methods have shown promising results in disease detection. Based on the concept features relevant to diseases, dictionary-based detection approaches and classical machine learning have shown effective performance (Rozova et al., 2023b; Martinez et al., 2015). Bag-of-words models have also been utilized, often combined with machine learning techniques to further enhance accuracy and scalability in disease detection (Cury et al., 2021; López-Úbeda et al., 2020). Recently, large language models (LLMs), such as BioBERT (Lee et al., 2020) and ClinicalBERT (Huang et al., 2019), pre-trained on large biomedical corpora, have improved contextual understanding in clinical texts (Consoli et al., 2024; Han et al., 2025).

Low resource settings remain challenging for NLP tasks. Few-shot fine-tuning (Brown et al., 2020; Gu et al., 2022; Liu et al., 2022), where large pre-trained models are adapted using only a small number of labeled examples, has shown promising results. Selecting effective few-shot samples is critical, and active learning strategies such as uncertainty sampling (Nguyen et al., 2022) and diversity sampling (Yang et al., 2015) are often employed. However, these methods typically optimize a single metric, and under domain shift, tend to select distributional outliers rather than truly informative samples (Gonsior et al., 2024). Reinforcement Learning (RL) (Fang et al., 2017; Liu et al., 2024b) offers a potential solution by optimizing more flexible and adaptive sample selection policies, thereby improving robustness in different contexts.

Class imbalance is especially crucial in low-resource clinical NLP tasks (Ghosh et al., 2024). Data-level approaches, such as oversampling minority classes (Hairani et al., 2024) and undersampling majority classes (Yang et al., 2024), are typically used to balance class distributions. Algorithm-level methods, such as cost-sensitive learning (Araf et al., 2024) and focal loss adjustments (Aljohani

et al., 2023), aim to direct model attention towards underrepresented classes, thereby improving model performance in class-imbalanced settings.

3 Methodology

3.1 Problem Setup and Overview

We study low-resource and class-imbalanced transfer learning between heterogeneous clinical report datasets: a fully labeled source dataset $\mathcal{D}_s$ and an unlabeled target dataset $\mathcal{U}_t$. Although the two datasets (domains) share some similar clinical knowledge, distribution shift and differences in label distribution make direct transfer challenging.

We formulate cross-domain adaptation as a budgeted active learning problem: given an annotation budget $B \ll N_t$, where N_t is the target pool size, our goal is to select a small but high-utility subset $\mathcal{Q} \subset \mathcal{U}_t$. The selected samples are then annotated and merged with $\mathcal{D}_s$ to form an expanded training set. With supervised fine-tuning on the final dataset, the knowledge can be effectively transferred and the model performance across both domains also improved.

The overall framework of RADS is shown in Figure 2. Our approach consists of three stages: (1) we train an active learner on $\mathcal{D}_s$ and compute informativeness signals for $\mathcal{U}_t$ via Monte-Carlo (MC) dropout; (2) we define a prior-aware utility that combines BALD-based mutual information (Houlsby et al., 2011) with pseudo-label class weighting to explicitly control the quality of selected samples for transfer learning under severe class imbalance; and (3) we train a reinforcement learning sampler to select samples that maximize prior-aware utility while discouraging redundant selections. The pseudocode for this part is provided in Appendix A.

3.2 Active Learner

We first fine-tune a lightweight classifier f_ϕ on the labeled source dataset $\mathcal{D}_s$. For each unlabeled target report in training pool $x \in \mathcal{U}_t$, we estimate epistemic uncertainty via MC dropout (Gal and Ghahramani, 2016). Specifically, we keep dropout activated at inference time and perform K stochastic forward passes. Each pass corresponds to sampling a dropout mask, yielding a sampled set of network weights $\mathbf{w}_k$ and a predictive distribution:

$$p_k(y \mid x) = \mathrm{softmax}(f_\phi(x; \mathbf{w}_k)) \quad (1)$$

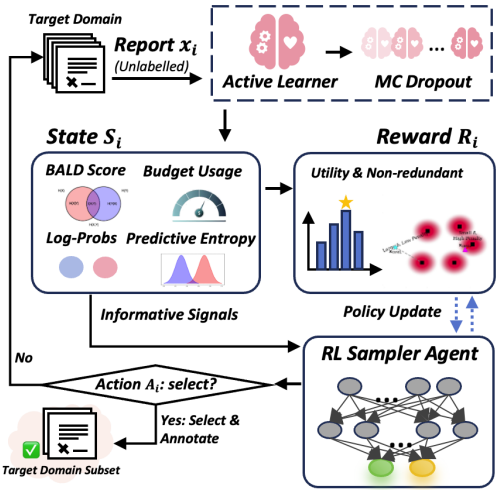

Figure 2: RADS framework for RL-based active sampling under domain shift. The active learner is fine-tuned on the source domain, and MC dropout is used to score unlabeled target reports and construct informativeness signals (state). An RL sampler then selects a subset for annotation by maximizing the reward, producing a target set for joint fine-tuning with the source data.

Aggregating these K stochastic predictions approximates the posterior predictive distribution. We compute the MC predictive mean as:

$$\bar{p}(y \mid x) = \frac{1}{K} \sum_{k=1}^{K} p_k(y \mid x) \quad (2)$$

In addition, we retain the mean log-probability vector $\bar{\ell}(x) = \log \bar{p}(\cdot \mid x)$, which serves as a representation for redundancy estimation in our RL-based sampler (Section 3.5).

Based on this active learner, we also define a pseudo label $\hat{y}(x) = \arg\max_y \bar{p}(y \mid x)$ and estimate the predicted target class prior:

$$\begin{aligned} \hat{\pi}_+ &= \frac{1}{N_t} \sum_{x \in \mathcal{U}_t} \mathbb{1}[\hat{y}(x) = 1], \\ \hat{\pi}_- &= 1 - \hat{\pi}_+. \end{aligned} \quad (3)$$

These priors let us correct selection bias when the pool is imbalanced or the source-trained model is miscalibrated on the target domain.

3.3 BALD Signal

To score informativeness for unlabeled target-domain samples, we use BALD, which quantifies the mutual information between the predicted label and the model parameters. Let $H(\cdot)$ denote entropy.

For each $x \in \mathcal{U}_t$, we compute:

$$\mathrm{PE}(x) = H(\bar{p}(\cdot \mid x)), \tag{4}$$

$$\mathrm{EE}(x) = \frac{1}{K}\sum_{k=1}^{K} H(p_k(\cdot \mid x)), \tag{5}$$

$$\mathrm{MI}(x) = \mathrm{PE}(x) - \mathrm{EE}(x). \tag{6}$$

Here, $\mathrm{MI}(x)$ is the BALD score. It is large when the predictive distribution is uncertain overall (high PE) while individual stochastic models are relatively confident but disagree with each other (low EE). We normalize $\mathrm{MI}(x)$ to $[0, 1]$ over $\mathcal{U}_t$, denoted as $\widetilde{\mathrm{MI}}(x)$.

We treat samples with high $\widetilde{\mathrm{MI}}(x)$ as informative and assign them higher utility in our selection policy. Prioritizing these samples for annotation is expected to reduce the model's uncertainty and improve transfer to the target domain.

3.4 Prior-Aware Utility for Sample Selection

Selecting the top-B uncertain samples can sometimes produce an extreme class skew. This often happens under domain shift and severe class imbalance, where the source-trained active learner may predict biased pseudo labels on the target domain. To control the selected class mixture, we introduce a prior-aware utility. We define class weights using the estimated prior:

$$\begin{aligned} w_+ &= \frac{\rho}{\mathrm{clip}(\hat{\pi}_+)}, \\ w_- &= \frac{1-\rho}{1-\mathrm{clip}(\hat{\pi}_+)}. \end{aligned} \tag{7}$$

where $\mathrm{clip}(\cdot)$ clamps probabilities away from $\{0, 1\}$ for stability and ρ is a hyperparameter that trades off class-balance control and informativeness. We then define the utility:

$$u(x) = \widetilde{\mathrm{MI}}(x) \cdot \begin{cases} w_+, & \hat{y}(x) = 1, \\ w_-, & \hat{y}(x) = 0. \end{cases} \tag{8}$$

This utility favors informative samples and shifts selection toward the desired class ratio.

3.5 RL-based Sample Selection Strategy

At each step t, the sampler agent observes the current candidate x_t and decides whether to select or discard. An episode ends when B samples are selected or the pool is exhausted.

State. For each candidate x_t, the state vector combines the active learner signals and a budget progress term:

$$s_t = \left[\, \bar{\ell}(x_t);\ \mathrm{PE}(x_t);\ \mathrm{MI}(x_t);\ |S_t|/B \,\right] \tag{9}$$

where $\bar{\ell}(x_t)$ is the mean log-probability vector computed from MC dropout; $\mathrm{PE}(x_t)$ is the predictive entropy; $\mathrm{MI}(x_t)$ is the BALD score; and $|S_t|/B$ indicates the fraction of the annotation budget already consumed, with S_t denoting the set of selected samples so far. In our binary setting, $\bar{\ell}(x_t) \in \mathbb{R}^2$, hence the overall dimension of the state vector is 5.

Reward. Our reward encourages the agent to select samples that are both (i) informative for learning under class imbalance and (ii) non-redundant with respect to previously selected instances. Specifically, when the agent selects the current candidate ($a_t = 1$) and the budget is not yet exhausted ($|S_t| < B$), we define:

$$r_t = u(x_t) - \lambda \cdot \mathrm{Red}(x_t, S_t) \tag{10}$$

and set $r_t = 0$ otherwise. Here, $u(x_t)$ is the prior-aware utility (Section 3.4) and λ controls the strength of the diversity regularization.

To discourage selecting near-duplicate samples, we measure redundancy in the active learner's predictive representation space. For a candidate x and the current selected set S, we first compute the distance to its nearest selected neighbor:

$$\delta(x, S) = \begin{cases} +\infty, & |S| = 0, \\ \min\limits_{x' \in S} \left\| \bar{\ell}(x) - \bar{\ell}(x') \right\|_2, & \text{otherwise.} \end{cases} \tag{11}$$

We then convert this distance into a bounded redundancy score:

$$\mathrm{Red}(x, S) = \begin{cases} 0, & |S| = 0, \\ \dfrac{1}{1+\delta(x,S)}, & \text{otherwise.} \end{cases} \tag{12}$$

This definition yields a larger penalty when x is very close to an existing selection (small δ), and a smaller penalty when x is far away (large δ). As a result, the agent is encouraged to select diverse samples while still prioritizing high-utility ones.

Dueling DQN Sampler Agent. We learn a Q-function $Q_\theta(s, a)$ with a dueling DQN architecture (Wang et al., 2016) and optimize it via the standard DQN objective (Mnih et al., 2015). We

Attribute	CHIFIR	PIFIR	MIMIC-CXR (subset)
Report Type	Cytology / Histopathology	PET–CT Radiology	Chest X-ray Radiology
Target Disease	Invasive Fungal Infection (IFI)	Invasive Fungal Infection (IFI)	Pneumonia
Label Type	Gold (manual)	Gold (manual)	Silver (auto-derived)
Class distribution (P/N)	14% / 86%	69% / 31%	39% / 61%
Dataset Size	283 reports (small)	201 reports (small)	493 reports (medium)

Table 1: Key attributes of the three datasets used in this study. P/N = positive/negative class proportions

maintain an experience replay buffer $\mathcal{B}$ and a target network Q_{θ^-}. At each gradient step, we minimize the temporal-difference loss:

$$\mathcal{L}(\theta) = \mathbb{E}_{(s,a,r,s',d)\sim\mathcal{B}}\Big[\big(Q_\theta(s,a) - y\big)^2\Big], \quad y = r + \gamma(1-d)\max_{a'} Q_{\theta^-}(s',a'). \tag{13}$$

where γ is the discount factor and $d \in \{0,1\}$ indicates episode termination. We adopt ϵ-greedy exploration with a decaying ϵ schedule, periodically synchronize θ^- with θ, and finally use the learned policy $\pi(s) = \arg\max_a Q_\theta(s,a)$ to select B samples from $\mathcal{U}_t$.

4 Experimental Setup

4.1 Benchmark Datasets

We chose three real world clinical datasets (Rozova et al., 2023a, 2025; Johnson et al., 2019) as benchmarks in this study: the PET-CT Invasive Fungal Infection Reports corpus (PIFIR[2]), the Cytology and Histopathology IFI Reports corpus (CHIFIR[3]), and the MIMIC Chest X-ray corpus (MIMIC-CXR[4]).

CHIFIR and PIFIR datasets are related to Invasive Fungal Infection (IFI), but the vocabulary used varies across them. The cytology and histopathology reports of the CHIFIR dataset assess tissue or fluid samples and describe the microscopic visualization of fungal organisms. The PET-CT reports from PIFIR assess metabolic activity and discuss the anatomical and morphological features of fungal lesions via PET imaging.

To assess transfer beyond IFI and beyond pathology-style reports, we also include MIMIC-CXR, a corpus of chest X-ray reports. We construct a *Pneumonia* subset by selecting the top 3,000 reports that are labeled as pneumonia by CheXpert's (Irvin et al., 2019) weak labels. Although PIFIR and MIMIC-CXR both consist of radiology reports, they still differ greatly in reporting style and clinical phrasing. Moreover, pneumonia and IFI reflect distinct clinical contexts, further increasing the domain shift. Figure 3 shows differences in predominant clinical terms across the three datasets.

All three datasets exhibit class imbalance. CHIFIR and MIMIC-CXR are dominated by negative cases, whereas PIFIR is dominated by positive cases. Throughout the paper we focus on transferring from other sources to PIFIR, and we provide results of other directions in the Appendix.

Table 1 summarizes the key characteristics of each dataset and highlights the challenges for transfer learning across them. Details of the dataset split are provided in Appendix B.

4.2 Evaluation Metrics

We evaluate performance using accuracy, F1 score, precision, recall, and ROC-AUC. Class imbalance in benchmark datasets makes the F1 score particularly important. Recall is also important, given that it is critical not to miss positive cases.

4.3 Baselines

We select the fine-tuned ClinicalBERT approach from previous work (Han et al., 2025) as the baseline. Table 2 shows the baseline results and reveals the challenges of knowledge transferability between these datasets. Models perform well when fine-tuned and evaluated on the same dataset. Without transfer learning, evaluation on a similar but still different dataset results in a clear performance drop. Although training on all datasets together can improve performance, it requires annotating all reports, which is labor-intensive. Full reproducibility

Figure 3: Word clouds for the CHIFIR (left), PIFIR (middle), and MIMIC-CXR (right) datasets. Word size corresponds to term frequency.

[2]Available for credentialed users at https://physionet.org/content/pifir/1.0.0/

[3]Available for credentialed users at https://physionet.org/content/corpus-fungal-infections/1.0.2/

[4]Available for credentialed users at https://physionet.org/content/mimic-cxr/2.1.0/

Transfer Learning to PIFIR Strategy	Datasets	Performance on PIFIR Acc	F1	P	R	Performance on CHIFIR Acc	F1	P	R	Performance on MIMIC-CXR Acc	F1	P	R
Baseline	PIFIR	0.714	0.812	0.788	0.839	–	–	–	–	–	–	–	–
Zero-shot	CHIFIR	0.357	0.229	1.000	0.129	0.942	0.824	0.778	0.875	–	–	–	–
	MIMIC-CXR	0.738	0.841	0.763	0.935	–	–	–	–	0.859	0.811	0.833	0.789
Full-shot	CHIFIR + PIFIR	0.857	0.900	0.931	0.871	0.904	0.615	0.800	0.500	–	–	–	–
	MIMIC-CXR + PIFIR	0.833	0.889	0.875	0.903	–	–	–	–	0.848	0.800	0.811	0.789

Table 2: Performance comparison of ClinicalBERT under different transfer strategies. Zero-shot transfer refers to fine-tuning solely on the source dataset. We set this as our baseline. Full-shot transfer refers to jointly fine-tuning on both source and target datasets. We assume this represents the best possible transfer learning performance.

details can be found in Appendix B.

5 Experimental Results

5.1 Transfer Learning Performance

We compare our method with several other active learning approaches to analyze the impact of different sample selection methods on knowledge transfer performance. 1) **Random Selection**: Randomly select samples from the unlabeled target domain. Each experiment is run five times to reduce variance and obtain more reliable results. We report the mean evaluation metrics over these five runs. 2) **Uncertainty-based Selection** (Nguyen et al., 2022): Selects k samples by predictive uncertainty (lowest confidence) from the active learner. 3) **Diversity-based Selection** (Yuan et al., 2020): Selects the k most diverse samples by calculating the cosine distance between each report embedding in the unlabeled dataset and the embeddings in the labeled dataset. 4) **LM-DPP Selection** (Wang et al., 2024): This method jointly models uncertainty and diversity using a Determinantal Point Process (DPP) kernel. Following the original work, we set the trade-off coefficient between uncertainty and diversity to 0.5 and select the subset of size k that maximizes the DPP objective for annotation. 5) **TAGCOS Selection** (Zhang et al., 2025): A task-agnostic selection baseline that selects k samples according to its gradient-based selection criterion. 6) **BatchBALD Selection** (Kirsch et al., 2019): Selects k samples using a batch acquisition strategy that extends BALD by maximizing joint mutual information under MC dropout.

Table 3 reports transfer learning results from CHIFIR to PIFIR. Uncertainty-, diversity-, and LM-DPP-based selection yield comparable or worse performance than random sampling. While TAGCOS and BatchBALD attain relatively high F1 scores on PIFIR, their ROC-AUC is noticeably lower. In contrast, our method RADS achieves the best performance on PIFIR while maintaining competitive performance on the source domain (CHIFIR). This indicates strong sample efficiency, requiring only $5/135 \approx 3.7\%$ of the target training set to obtain substantial transfer gains.

We further compare RADS with a prompt-guided LLM selection baseline (Jeong et al., 2025), which uses an open-source medical LLM to score report and then selects the top-k reports for annotation. We report the CHIFIR to PIFIR transfer results in Appendix C. Although this baseline can retrieve some useful reports as the budget increases, it is highly unstable in the ultra-low-budget regime and remains less reliable than RADS overall.

Table 4 reports transfer learning results from MIMIC-CXR to PIFIR. Other baselines provide limited gains and are often on par with or below random selection. RADS achieves better target performance (F1 on PIFIR = 0.882) and remains highly sample-efficient, requiring annotation of only $2/135 \approx 1.5\%$ of the target training set.

We also conduct transfer learning experiments from PIFIR to CHIFIR. The results show that our method still achieves the best performance with only 8 samples selected from target dataset CHIFIR. More detailed discussion is shown in Appendix D.

RADS consistently outperforms strong baselines, demonstrating superior sample-efficient transfer performance. More robustness analysis under imbalanced settings appears in Appendix E.

5.2 Learning Curves under Varying Budgets

We analyze the effect of annotation budget on transfer performance. Figure 4 shows the transfer performance from CHIFIR to PIFIR across budgets under two strong baselines. BatchBALD is highly unstable; a small change in budget can flip the model from almost perfect to completely broken. TAGCOS is more stable but still remains unreliable at small budgets. Figure 5 shows the transfer per-

Knowledge Transfer from CHIFIR to PIFIR												
Strategy	**Performance on PIFIR**					**Performance on CHIFIR**					**Transfer Gap**	
	Accuracy	F1-score	Precision	Recall	ROC-AUC	Accuracy	F1-score	Precision	Recall	ROC-AUC	ΔF1	95% CI
Random	0.595	0.639	0.885	0.561	0.813	0.927	0.746	0.805	0.700	0.938	–	–
Uncertainty	0.524	0.545	0.923	0.387	0.830	0.942	0.824	0.778	0.875	0.977	0.278	[-0.037, 0.530]
Diversity	0.595	0.638	0.938	0.484	0.809	0.942	0.800	0.857	0.750	0.974	0.162	[-0.167, 0.412]
LM-DPP	0.571	0.609	0.933	0.452	0.839	0.904	0.615	0.800	0.500	0.977	0.007	[-0.418, 0.319]
TAGCOS	0.762	0.844	0.818	0.871	0.730	0.942	0.824	0.778	0.875	0.972	-0.020	[-0.310, 0.168]
BatchBALD	0.738	0.849	0.738	1.000	0.783	0.885	0.500	0.750	0.375	0.946	-0.349	[-0.833, -0.019]
RADS	**0.810**	**0.871**	**0.871**	**0.871**	**0.833**	**0.923**	**0.750**	**0.750**	**0.750**	**0.977**	**-0.121**	**[-0.430, 0.100]**

Table 3: Transfer learning performance from CHIFIR to PIFIR with 5 samples selected in PIFIR under different sample selection strategies. ΔF1 = F1(CHIFIR) − F1(PIFIR). CI = Confidence Interval.

Knowledge Transfer from MIMIC-CXR to PIFIR												
Strategy	**Performance on PIFIR**					**Performance on MIMIC-CXR**					**Transfer Gap**	
	Accuracy	F1-score	Precision	Recall	ROC-AUC	Accuracy	F1-score	Precision	Recall	ROC-AUC	ΔF1	95% CI
Random	0.766	0.848	0.808	0.842	0.806	0.862	0.824	0.808	0.842	0.932	–	–
Uncertainty	0.738	0.831	0.794	0.871	0.636	0.848	0.810	0.780	0.842	0.916	-0.021	[-0.165, 0.126]
Diversity	0.738	0.845	0.750	0.968	0.616	0.859	0.816	0.816	0.816	0.933	-0.029	[-0.157, 0.102]
LM-DPP	0.738	0.831	0.794	0.871	0.636	0.848	0.810	0.780	0.842	0.916	-0.021	[-0.165, 0.126]
TAGCOS	0.738	0.836	0.778	0.903	0.795	0.862	0.831	0.821	0.842	0.930	-0.005	[-0.131, 0.139]
BatchBALD	0.762	0.848	0.800	0.903	0.827	0.889	0.861	0.829	0.895	0.949	0.012	[-0.107, 0.141]
RADS	**0.810**	**0.882**	**0.811**	**0.968**	**0.880**	**0.869**	**0.840**	**0.791**	**0.895**	**0.921**	**-0.043**	**[-0.153, 0.072]**

Table 4: Transfer learning performance from MIMIC-CXR to PIFIR with 2 samples selected in PIFIR under different sample selection strategies. ΔF1 = F1(MIMIC-CXR) − F1(PIFIR). CI = Confidence Interval.

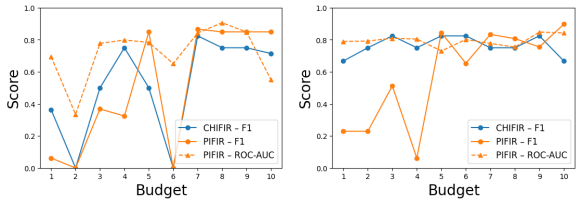

Figure 4: Transfer from CHIFIR to PIFIR under baselines BatchBALD (left) and TAGCOS (right).

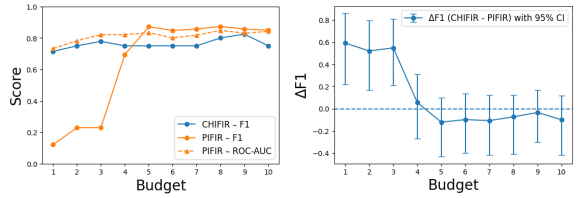

Figure 5: Transfer from CHIFIR to PIFIR under our method RADS.

formance from CHIFIR to PIFIR with our method. The left graph shows that starting from budget 5, F1 on PIFIR stays around 0.85–0.87 and additional labels provide marginal improvements, while CHIFIR performance remains stable. The right graph plots the domain gap ΔF1 versus budget with 95% confidence intervals. From budget = 4 onward, the gap is effectively closed (ΔF1 ≈ 0 with overlapping confidence intervals), suggesting the selected subset suffices to eliminate the transfer gap.

As MIMIC-CXR is larger than CHIFIR, identifying informative target samples is less challenging for all models, even under low annotation budgets. The transfer performance from MIMIC-CXR to PIFIR across budgets is shown in Appendix F.

5.3 Ablation Study

Method	**Accuracy**	**F1-score**	**Precision**	**Recall**	**ROC-AUC**
No RL	0.571	0.609	0.933	0.452	0.798
MI Only	0.262	0.000	0.000	0.000	0.475
Utility Only	0.786	0.866	0.806	0.935	0.833
RADS	**0.810**	**0.871**	**0.871**	**0.871**	**0.833**

Table 5: Ablation results under the hard transfer setting from CHIFIR to PIFIR with a labeling budget of 5.

To evaluate the effectiveness of each component in RADS, we conduct ablation studies as shown in Table 5. Replacing the RL sampler with a greedy selector (No RL) leads to a clear drop performance. Although this variant optimizes the same objective, it lacks the sequential decision-making needed to balance exploration and redundancy control. Selecting samples by BALD signal (MI Only) fails, implying that uncertainty-only criteria can favor noisy or out-of-distribution target examples under domain shift. Using the prior-aware utility (Utility Only) greatly improves results, confirming the benefit of class- and quality-aware selection, but remains below our method, highlighting the ad-

ditional gains from discouraging redundant selections. Finally, RADS further improves over Utility Only, demonstrating that the RL sampler provides additional gains by learning a non-redundant, globally optimized subset selection policy rather than relying on pointwise ranking.

5.4 Selected Sample Quality Analysis

We audit the quality of the selected target samples. Because RADS is trained and applied on the same unlabeled target pool, we clarify that the sampler never accesses target gold labels during optimization. Figure 6 shows that in CHIFIR to PIFIR transfer, the source-trained pseudo labels are notably misaligned with the annotated labels, yet RADS still selects mostly true positives, which helps improve target performance. In MIMIC-CXR to PIFIR, the pseudo-label ratio better matches the annotated composition, consistent with smaller domain shift and improved calibration on the target domain.

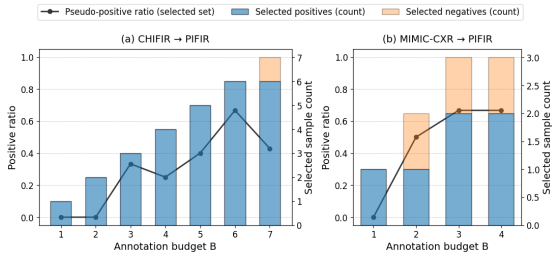

Figure 6: Selected sample analysis under CHIFIR to PIFIR (left) and MIMIC-CXR to PIFIR (right) transfer. The black line shows the pseudo-positive ratio predicted by the source-trained active learner, and the bars report the numbers of true positives and true negatives after manual annotation (blue/yellow).

Figure 7 visualizes the CHIFIR to PIFIR (left) and MIMIC-CXR to PIFIR (right) transfer, where our method selects 5 samples for adaptation. Before transfer learning, the decision boundary only captures the source dataset specific separation. After adding the selected target subset, the boundary rotates and shifts toward a direction that better reflects the class layout from both datasets.

5.5 Efficiency and Budget-Aware Transfer

Runtime Efficiency Our RL-based sampler in RADS is light-weight, requiring only a few seconds to produce a selection. Selecting 2 samples (MIMIC-CXR to PIFIR transfer) takes around 3 seconds, and selecting 5 samples (CHIFIR to PIFIR transfer) takes around 9 seconds. Given that

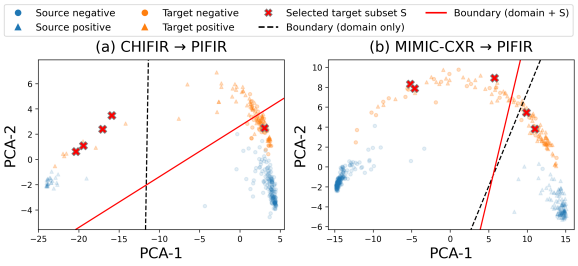

Figure 7: Transfer learning from CHIFIR to PIFIR (left) and MIMIC-CXR to PIFIR (right) with PCA projection of report embeddings. Red $\times$ markers denote the selected PIFIR subset S. The dashed black line shows the decision boundary learned from the source dataset only. The solid red line shows the boundary after augmenting with selected samples. Source = CHIFIR / MIMIC-CXR dataset, Target = PIFIR datatset.

annotating a single report takes about one minute, the selection overhead is negligible. Compared with full-shot transfer that labels the entire target training set (135 reports), RADS achieves comparable target performance with only 2 or 5 annotated reports.

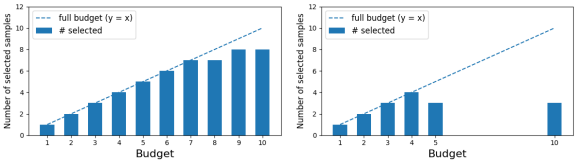

Figure 8: Number of selected PIFIR samples versus the annotation budget B when training on CHIFIR to PIFIR (left) and MIMIC-CXR to PIFIR (right).

Budget Utilization and Early Stopping As our RL-based sampler encodes budget progress ($|S_t|/B$) in the state, it can also provide guidance on how many samples are worth annotating during transfer. Figure 8 shows the actual number of PIFIR samples selected as the budget increases. For CHIFIR to PIFIR transfer, the policy stops fully consuming the budget once $B \geq 8$, consistent with our results showing good transfer performance already at $B = 5$. For MIMIC-CXR to PIFIR transfer, the policy no longer uses the full budget when $B \geq 5$, aligning with good target performance achieved with $B = 2$ labeled PIFIR samples. This pattern is useful as in active learning it is important to know when it is time to stop adding samples.

5.6 Transfer Gap between Datasets

To explain why model transfer from MIMIC-CXR to PIFIR is easier than CHIFIR to PIFIR, we further

analyze the distribution gap between these datasets.

We quantify overlap between datasets with a shared unigram–bigram vocabulary (Elangovan et al., 2021). Coverage of PIFIR-test n-grams is higher for MIMIC-CXR to PIFIR than for CHIFIR to PIFIR (0.193 vs. 0.115). The Jaccard similarity between source and target vocabularies is also higher for MIMIC-CXR to PIFIR than for CHIFIR to PIFIR (0.187 vs. 0.124). This suggests a smaller lexical domain shift between MIMIC-CXR and PIFIR, as it is illustrated by our empirical results fine-tuning MIMIC-CXR. More detailed differences are analyzed in Appendix G.

6 Conclusion

In this work, we studied transfer learning for disease detection under low-resource and class-imbalanced conditions. We proposed RADS, an RL-based sampler that jointly optimizes a prior-aware utility for class-mixture control and a diversity regularizer to avoid near-duplicate selections. Our approach improves model performance and adaptability across medical datasets compared to traditional sample selection strategies. We expect this approach to generalize to other transfer learning problems not only in clinical NLP but also in broader application domains, offering a promising direction for broader validation and impact.

Limitations

Despite demonstrating promising results, our approach has several limitations. First, the effectiveness of our RL-based sample selection heavily depends on the feedback provided by the active learner. This places high quality demands on the original gold dataset. Second, our formulation controls class mixture only through predicted priors and does not explicitly incorporate richer clinical knowledge. Third, our experiments focus on binary clinical disease detection with relatively small target pools and we have not yet validated RADS on larger-scale multi-class settings or on broader non-clinical transfer tasks. Fourth, more stable RL optimizers, improved uncertainty estimation, and better transfer-aligned validation strategies remain important directions for future work.

Acknowledgments

This work was supported by the Australian Government through Medical Research Future Fund grant MRFCRI000188. The authors also thank the EINSTEIN Study Group for their support and insightful discussions.

References

Naif Radi Aljohani, Ayman Fayoumi, and Saeed-Ul Hassan. 2023. A novel focal-loss and class-weight-aware convolutional neural network for the classification of in-text citations. *Journal of Information Science*, 49(1):79–92.

Zaid Alyafeai, Maged Saeed AlShaibani, and Irfan Ahmad. 2020. A survey on transfer learning in natural language processing. *arXiv preprint arXiv:2007.04239*.

Imane Araf, Ali Idri, and Ikram Chairi. 2024. Cost-sensitive learning for imbalanced medical data: a review. *Artificial Intelligence Review*, 57(4):80.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, and 12 others. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Sergio Consoli, Peter Markov, Nikolaos I Stilianakis, Lorenzo Bertolini, Antonio Puertas Gallardo, and Mario Ceresa. 2024. Epidemic information extraction for event-based surveillance using large language models. In *International Congress on Information and Communication Technology*, pages 241–252. Springer Nature Singapore Singapore.

Ricardo C Cury, Istvan Megyeri, Tony Lindsey, Robson Macedo, Juan Batlle, Shwan Kim, Brian Baker, Robert Harris, and Reese H Clark. 2021. Natural Language Processing and Machine Learning for Detection of Respiratory Illness by Chest CT Imaging and Tracking of COVID-19 Pandemic in the United States. *Radiology: Cardiothoracic Imaging*, 3(1):e200596.

Aparna Elangovan, Jiayuan He, and Karin Verspoor. 2021. Memorization vs. generalization : Quantifying data leakage in NLP performance evaluation. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1325–1335, Online. Association for Computational Linguistics.

Meng Fang, Yuan Li, and Trevor Cohn. 2017. Learning how to active learn: A deep reinforcement learning approach. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 595–605, Copenhagen, Denmark. Association for Computational Linguistics.

Yifan Fu, Xingquan Zhu, and Bin Li. 2013. A survey on instance selection for active learning. *Knowledge and information systems*, 35(2):249–283.

Yarin Gal and Zoubin Ghahramani. 2016. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR.

Kushankur Ghosh, Colin Bellinger, Roberto Corizzo, Paula Branco, Bartosz Krawczyk, and Nathalie Japkowicz. 2024. The class imbalance problem in deep learning. *Machine Learning*, 113(7):4845–4901.

Julius Gonsior, Christian Falkenberg, Silvio Magino, Anja Reusch, Claudio Hartmann, Maik Thiele, and Wolfgang Lehner. 2024. Comparing and improving active learning uncertainty measures for transformer models by discarding outliers. *Information systems frontiers*, pages 1–17.

Yuxian Gu, Xu Han, Zhiyuan Liu, and Minlie Huang. 2022. PPT: Pre-trained prompt tuning for few-shot learning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8410–8423, Dublin, Ireland. Association for Computational Linguistics.

Hairani Hairani, Triyanna Widiyaningtyas, and Didik Dwi Prasetya. 2024. Addressing class imbalance of health data: A systematic literature review on modified synthetic minority oversampling technique (smote) strategies. *JOIV: International Journal on Informatics Visualization*, 8(3):1310–1318.

Wei Han, David Martinez, Vlada Rozova, Lawrence Cavedon, Anna Khanina, Leon J Worth, Monica A Slavin, Karin A Thursky, and Karin Verspoor. 2025. Automated detection of invasive fungal infections in clinical reports using medical language models. *Studies in Health Technology and Informatics*, 329:1002–1007.

Neil Houlsby, Ferenc Huszár, Zoubin Ghahramani, and Máté Lengyel. 2011. Bayesian active learning for classification and preference learning. *arXiv preprint arXiv:1112.5745*.

Kexin Huang, Jaan Altosaar, and Rajesh Ranganath. 2019. ClinicalBERT: Modeling clinical notes and predicting hospital readmission. *CHIL 2020 Workshop*.

Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn Ball, Katie Shpanskaya, and 1 others. 2019. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 590–597.

Henrik Elvang Jensen. 2021. Histopathology in the diagnosis of invasive fungal diseases. *Current Fungal Infection Reports*, 15(1):23–31.

Daniel P Jeong, Zachary Chase Lipton, and Pradeep Kumar Ravikumar. 2025. LLM-select: Feature selection with large language models. *Transactions on Machine Learning Research*.

Alistair EW Johnson, Tom J Pollard, Seth J Berkowitz, Nathaniel R Greenbaum, Matthew P Lungren, Chih-ying Deng, Roger G Mark, and Steven Horng. 2019. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Scientific data*, 6(1):317.

Justin M Johnson and Taghi M Khoshgoftaar. 2019. Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1):1–54.

Andreas Kirsch, Joost Van Amersfoort, and Yarin Gal. 2019. Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning. *Advances in neural information processing systems*, 32.

Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2020. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240.

Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, and 1 others. 2024a. DeepSeek-V3 Technical Report. *arXiv preprint arXiv:2412.19437*.

Haokun Liu, Derek Tam, Mohammed Muqeeth, Jay Mohta, Tenghao Huang, Mohit Bansal, and Colin A Raffel. 2022. Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning. *Advances in Neural Information Processing Systems*, 35:1950–1965.

Ying Liu, Haozhu Wang, Huixue Zhou, Mingchen Li, Yu Hou, Sicheng Zhou, Fang Wang, Rama Hoetzlein, and Rui Zhang. 2024b. A review of reinforcement learning for natural language processing and applications in healthcare. *Journal of the American Medical Informatics Association*, 31(10):2379–2393.

Pilar López-Úbeda, Manuel Carlos Díaz-Galiano, Teodoro Martín-Noguerol, Antonio Luna, L Alfonso Ureña-López, and M Teresa Martín-Valdivia. 2020. COVID-19 detection in radiological text reports integrating entity recognition. *Computers in Biology and Medicine*, 127:104066.

Mieradilijiang Maimaiti, Yang Liu, Huanbo Luan, and Maosong Sun. 2021. Enriching the transfer learning with pre-trained lexicon embedding for low-resource neural machine translation. *Tsinghua Science and Technology*, 27(1):150–163.

David Martinez, Michelle R Ananda-Rajah, Hanna Suominen, Monica A Slavin, Karin A Thursky, and Lawrence Cavedon. 2015. Automatic detection of patients with invasive fungal disease from free-text computed tomography (CT) scans. *Journal of Biomedical Informatics*, 53:251–260.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, and 1 others. 2015. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533.

Vu-Linh Nguyen, Mohammad Hossein Shaker, and Eyke Hüllermeier. 2022. How to measure uncertainty in uncertainty sampling for active learning. *Machine Learning*, 111(1):89–122.

Vlada Rozova, Anna Khanina, Jeremy Ong, Ramin Alipour, Leon Worth, Monica Slavin, Karin Thursky, and Karin Verspoor. 2025. PIFIR: PET-CT Invasive Fungal Infection Reports. *PhysioNet*. Version 1.0.0.

Vlada Rozova, Anna Khanina, Jasmine Teng, Joanne Teh, Leon Worth, Monica Slavin, karin thursky, and Karin Verspoor. 2023a. CHIFIR: Cytology and Histopathology Invasive Fungal Infection Reports. *PhysioNet*. Version 1.0.0.

Vlada Rozova, Anna Khanina, Jasmine C Teng, Joanne SK Teh, Leon J Worth, Monica A Slavin, Karin A Thursky, and Karin Verspoor. 2023b. Detecting evidence of invasive fungal infections in cytology and histopathology reports enriched with concept-level annotations. *Journal of Biomedical Informatics*, 139:104293.

Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, and Chunfang Liu. 2018. A survey on deep transfer learning. In *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27*, pages 270–279. Springer.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, and 1 others. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

David W Townsend, Jonathan PJ Carney, Jeffrey T Yap, and Nathan C Hall. 2004. PET/CT today and tomorrow. *Journal of Nuclear Medicine*, 45(1 suppl):4S–14S.

Peng Wang, Xiaobin Wang, Chao Lou, Shengyu Mao, Pengjun Xie, and Yong Jiang. 2024. Effective demonstration annotation for in-context learning via language model-based determinantal point process. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 1266–1280, Miami, Florida, USA. Association for Computational Linguistics.

Ziyu Wang, Tom Schaul, Matteo Hessel, Hado Hasselt, Marc Lanctot, and Nando Freitas. 2016. Dueling network architectures for deep reinforcement learning. In *International conference on machine learning*, pages 1995–2003. PMLR.

Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. 2016. A survey of transfer learning. *Journal of Big Data*, 3:1–40.

Cynthia Yang, Egill A Fridgeirsson, Jan A Kors, Jenna M Reps, and Peter R Rijnbeek. 2024. Impact of random oversampling and random undersampling on the performance of prediction models developed using observational health data. *Journal of Big Data*, 11(1):7.

Yi Yang, Zhigang Ma, Feiping Nie, Xiaojun Chang, and Alexander G Hauptmann. 2015. Multi-class active learning by uncertainty sampling with diversity maximization. *International Journal of Computer Vision*, 113:113–127.

Michelle Yuan, Hsuan-Tien Lin, and Jordan Boyd-Graber. 2020. Cold-start active learning through self-supervised language modeling. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7935–7948, Online. Association for Computational Linguistics.

Jipeng Zhang, Yaxuan Qin, Renjie Pi, Weizhong Zhang, Rui Pan, and Tong Zhang. 2025. TAGCOS: Task-agnostic gradient clustered coreset selection for instruction tuning data. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 4671–4686.

A Algorithm Pseudocode

We provide the pseudocode for the strategy of sample selection in our approach below.

Require: pool $\mathcal{U}_t$, budget B, episodes N, utility $u(\cdot)$, diversity weight λ
feature set $\mathcal{F} = \{\bar{\ell}(x), \mathrm{PE}(x), \widetilde{\mathrm{MI}}(x), |S|/B\}$, action set $\mathcal{A} = \{0, 1\}$
env $\leftarrow$ RLSAMPLESELECTIONENV($\mathcal{U}_t, \mathcal{F}, u, B, \lambda$)
Initialize online network Q_ϕ and target network $\hat{Q}_\phi \leftarrow Q_\phi$
Initialize replay buffer $\mathcal{D} \leftarrow \emptyset$
Initialize exploration rate ϵ
for *episode* $= 1$ **to** N **do**
  $s \leftarrow$ *env.reset()*; *done* $\leftarrow$ **false**
  while not *done* **do**
    $a \leftarrow$ EPSGREEDY(Q_ϕ, s, ϵ)
    $(s', r,$ *done*$) \leftarrow$ *env.step*(a)
    Add $(s, a, r, s',$ *done*$)$ to $\mathcal{D}$
    if $|\mathcal{D}| \geq M$ **then** ▷ M is minibatch size
      $\mathcal{M} \leftarrow$ SAMPLEMINIBATCH($\mathcal{D}, M$)
      UPDATENETS($Q_\phi, \hat{Q}_\phi, \mathcal{M}$)
    end if
    $s \leftarrow s'$
  end while
  if *episode* $\bmod K_{\text{upd}} = 0$ **then**
    $\hat{Q}_\phi \leftarrow Q_\phi$
  end if
  $\epsilon \leftarrow$ DECAYEPS(ϵ)
end for
Selection (Greedy Policy)
$\mathcal{S} \leftarrow \emptyset$; $s \leftarrow$ *env.reset()*; *done* $\leftarrow$ **false**
while not *done* **do**
  id $\leftarrow$ *env.currentId()*
  $a \leftarrow \arg\max_{a' \in \mathcal{A}} Q_\phi(s, a')$
  $(s',$ _, *done*$) \leftarrow$ *env.step*(a)
  if $a = 1$ **then**
    $\mathcal{S} \leftarrow \mathcal{S} \cup \{id\}$
  end if
  $s \leftarrow s'$
end while
return $\mathcal{S}$

B Reproducibility

All experiments are conducted on a single NVIDIA A100 GPU. Key fine-tuning settings are: epochs = 15, learning rate = 2×10^{-5}, batch size = 8, max sequence length = 512, weight decay = 0.01, and early stopping with a patience of 3 epochs.

For uncertainty estimation, we use MC dropout with the number of stochastic forward passes set to $K = 10$. In RADS, we train a dueling DQN sampler for 300 episodes with ϵ-greedy exploration, decaying ϵ from 1.0 to 0.05 with a multiplicative factor of 0.995. We use an experience replay buffer of size 10000 and start network updates once at least one minibatch is available (batch size = 64). Both the online and target networks are optimized with Adam (learning rate = 10^{-4}) and discount factor $\gamma = 0.95$, and the target network is synchronized every 10 episodes. $\rho = 0.9$. The reward is defined as the prior-aware utility minus a diversity penalty computed from the nearest-neighbor distance in the mean log-probability space $\bar{\ell}(x)$, with $\lambda = 0.01$.

Dataset Split Each dataset is split into training, validation, and test sets (around 70%, 10%, and 20%), preserving the original class balance. Table 8 shows the number of positive and negative samples in each split.

C Prompt-guided LLM Selection Baseline

We compare RADS against a prompt-guided LLM selection baseline inspired by recent LLM-based data selection methods (Jeong et al., 2025). In this baseline, we use an open-source medical LLM (OpenBioLLM-8B[5]) to score each unlabeled target report according to its estimated usefulness for training the IFI classifier, and then select the top-k reports for annotation.

Table 9 reports results for transfer from CHIFIR to PIFIR under different annotation budgets. We observe that the LLM-guided baseline is highly unstable in the ultra-low-budget regime: at budgets 1 and 4, it completely fails to recover positive target performance, and at budget 2 it yields only a marginal F1 improvement. Performance improves at larger budgets, suggesting that the LLM can identify some useful reports when more annotations are allowed. However, the overall behavior remains much less reliable than RADS in the low-budget regime that is central to our study.

D Transfer Performance from PIFIR to CHIFIR

Table 6 summarizes the baseline transfer performance. The zero-shot model trained on PIFIR exhibits severe performance degradation on CHIFIR, achieving very low accuracy and F1, indicating a substantial domain shift. In particular, the model tends to over-predict the positive class on CHIFIR (high recall but very low precision), which suggests that the decision boundary learned from PIFIR does not directly generalize to CHIFIR. When CHIFIR data are available for adaptation (full-shot), incorporating CHIFIR supervision mitigates this shift and improves target-domain performance, demon-

[5] https://huggingface.co/aaditya/Llama3-OpenBioLLM-8B

Transfer Learning to CHIFIR Strategy	Datasets	Performance on CHIFIR Acc	F1	P	R	Performance on PIFIR Acc	F1	P	R
Baseline	CHIFIR	0.942	0.824	0.778	0.875	–	–	–	–
Zero-shot	PIFIR	0.154	0.267	0.154	1.000	0.714	0.812	0.788	0.839
Full-shot	PIFIR + CHIFIR	0.904	0.615	0.800	0.500	0.857	0.900	0.931	0.871

Table 6: Performance comparison of transfer learning from PIFIR to CHIFIR under zero-shot transfer and full-shot transfer.

Knowledge Transfer from PIFIR to CHIFIR												
Strategy	Performance on CHIFIR					Performance on PIFIR					Transfer Gap	
	Accuracy	F1-score	Precision	Recall	ROC-AUC	Accuracy	F1-score	Precision	Recall	ROC-AUC	ΔF1	95% CI
Random	0.838	0.167	0.250	0.125	0.716	0.752	0.791	0.916	0.729	0.837	–	–
Uncertainty	0.788	0.267	0.286	0.250	0.656	0.905	0.938	0.909	0.968	0.880	0.671	[0.387, 0.967]
Diversity	0.154	0.267	0.154	1.000	0.591	0.738	0.849	0.738	1.000	0.883	0.584	[0.409, 0.752]
LM-DPP	0.154	0.267	0.154	1.000	0.489	0.738	0.849	0.738	1.000	0.815	0.583	[0.409, 0.752]
TAGCOS	0.769	0.143	0.167	0.125	0.545	0.929	0.954	0.912	1.000	0.827	0.811	[0.529, 0.986]
BatchBALD	0.846	0.000	0.000	0.000	0.486	0.714	0.793	0.852	0.742	0.742	0.793	[0.654, 0.897]
RADS	**0.865**	**0.632**	**0.545**	**0.750**	**0.858**	**0.881**	**0.921**	**0.906**	**0.935**	**0.865**	**0.289**	**[0.075, 0.599]**

Table 7: Transfer learning performance from PIFIR to CHIFIR with 8 samples selected in CHIFIR under different sample selection strategies. ΔF1 = F1(PIFIR) − F1(CHIFIR). CI = Confidence Interval.

Split	CHIFIR Total	P	N	MIMIC-CXR Total	P	N	PIFIR Total	P	N
Train	196	27	169	320	124	196	135	92	43
Dev	35	5	30	74	29	45	24	16	8
Test	52	8	44	99	38	61	42	31	11

Table 8: Class distribution for CHIFIR, MIMIC-CXR, and PIFIR across train, development, and test sets. P = the number of positive reports, N = the number of negative reports.

Num	Accuracy	F1-score	Precision	Recall
1	0.261	0.000	0.000	0.000
2	0.285	0.062	1.000	0.032
4	0.261	0.000	0.000	0.000
8	0.642	0.681	1.000	0.516
16	0.571	0.591	1.000	0.419
32	0.942	0.800	0.857	0.750

Table 9: Prompt-guided LLM selection results for CHIFIR to PIFIR transfer. Num is the number of selected PIFIR reports.

strating that a small amount of target data is critical for reliable transfer.

We next evaluate whether our data selection strategy can maximize the benefit of limited target supervision. Table 7 reports transfer learning results when only 8 CHIFIR samples are selected for fine-tuning under different sampling strategies. Across all baselines, we observe that other strategies are insufficient to bridge the transfer gap: they either fail to improve CHIFIR F1 or produce unstable behavior. In contrast, RADS achieves the strongest transfer performance on CHIFIR, yielding the best overall target metrics (Accuracy/F1/ROC-AUC) while maintaining high performance on the source domain. Importantly, RADS also produces the smallest transfer gap (ΔF1) among compared methods, indicating that the selected CHIFIR samples lead to more effective adaptation without sacrificing the knowledge learned from PIFIR. The confidence interval of ΔF1 further suggests that RADS provides a more reliable and stable reduction of the transfer discrepancy compared to alternative selection strategies.

E Robustness under Imbalanced Sampling

We evaluate the robustness of our sampling strategy under imbalanced sampling. For transfer learning from CHIFIR to PIFIR, We randomly select 5 samples from PIFIR with different positive to negative ratios. Each setting is repeated five times to obtain stable results. Figure 10 shows the results. The best performance occurs when the positive to negative ratio is 1.00:0.00, which matches the class ratio selected by our method. This happens because CHIFIR has many more negative cases. Prioritizing positive target samples helps counter this imbalance and narrow the target-domain class distribution gap during transfer.

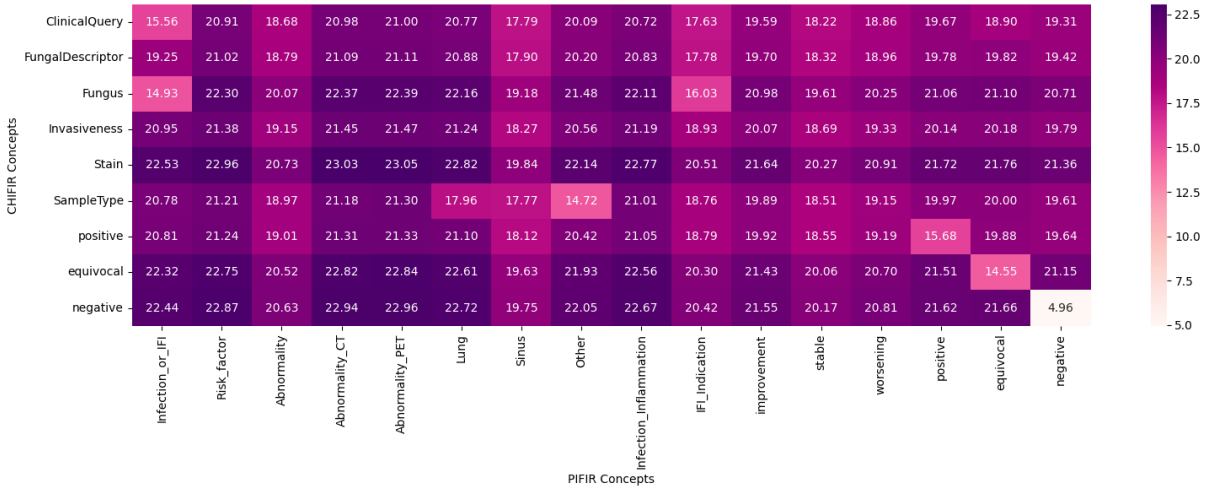

Figure 9: Concept-level KL divergence from CHIFIR to PIFIR.

CHIFIR		PIFIR		MIMIC-CXR	
Word	**TF-IDF**	**Word**	**TF-IDF**	**Word**	**TF-IDF**
cells	18.405196	uptake	17.527355	chest	33.730506
fluid	12.311106	ct	16.972962	pneumonia	30.779783
bronchial	11.999560	fdg	14.968679	right	27.327142
description	11.998803	pet	11.530773	left	24.813217
biopsy	11.490593	right	9.951929	pleural	24.584867
tissue	10.625458	marrow	8.764460	effusion	22.224288
specimen	9.652244	activity	8.720572	pulmonary	21.700909
lung	9.026040	disease	8.699531	lung	21.272838
washings	9.022390	left	8.493931	comparison	21.251147
cell	8.798129	findings	8.308830	findings	20.950609

Table 10: Top 10 terms with the highest TF–IDF scores in CHIFIR, PIFIR and MIMIC-CXR.

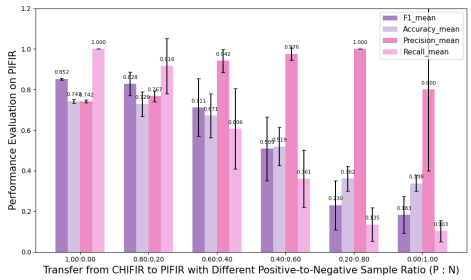

Figure 10: Class imbalance analysis of positive to negative sample ratios for CHIFIR to PIFIR transfer. Bars show mean values and black lines indicate variance.

F Learning Curves for MIMIC-CXR to CHIFIR Transfer under Varying Budgets

Figure 12 shows the performance from MIMIC-CXR to PIFIR across budgets under two baselines and Figure 13 (left) shows our methods' performance. MIMIC-CXR is larger and the zero-shot baseline is already good. Therefore, the headroom for improvement is limited, and our method is similar to other baselines. Figure 13 (right) plots the domain gap ΔF1 against budget with 95% confidence intervals. Across budgets, the point estimates stay close to zero and the confidence intervals largely overlap, indicating a small residual gap and no clear separation between budgets in this ultra-low-resource setting.

G Differences Analysis in CHIFIR, PIFIR, and MIMIC-CXR Datasets

CHIFIR contains 283 reports from 201 patients, with an average report length of 1,353 characters. PIFIR contains 201 reports from 156 patients, with an average report length of 1,809 characters. The MIMIC-CXR subset contains 493 reports from 290 patients, with a shorter average report length of 677 characters.

We first compute TF–IDF scores within each corpus and compare the top 10 highest-scoring terms between CHIFIR, PIFIR and MIMIC-CXR as shown in Table 10. CHIFIR is dominated by pathology and specimen-centric language (e.g.,

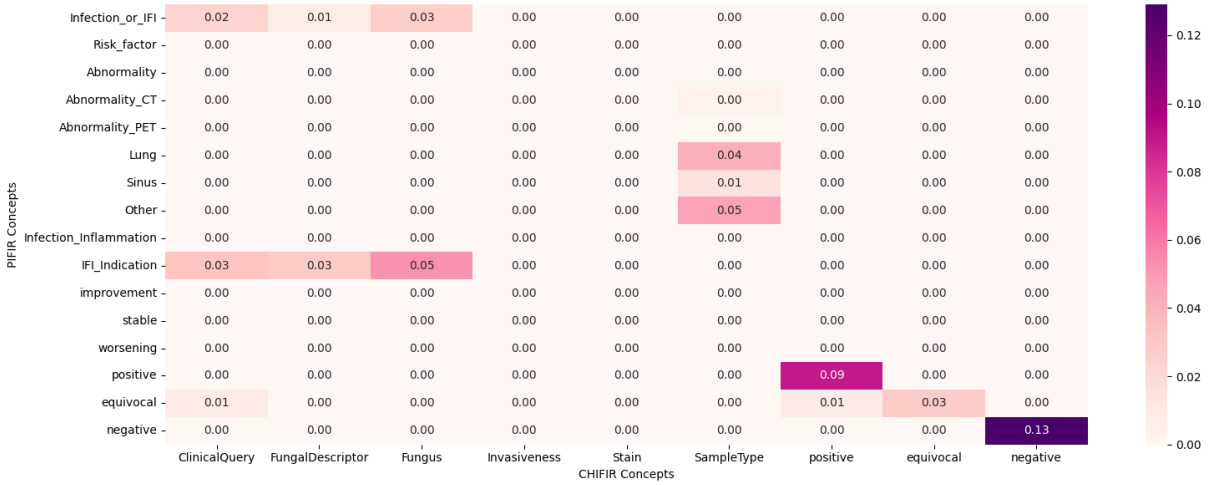

Figure 11: Jaccard similarity heatmap between CHIFIR and PIFIR concepts.

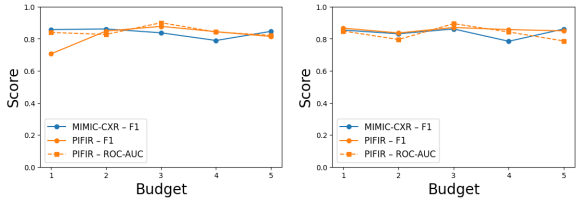

Figure 12: Transfer from MIMIC-CXR to PIFIR under baselines BatchBALD (left) and TAGCOS (right).

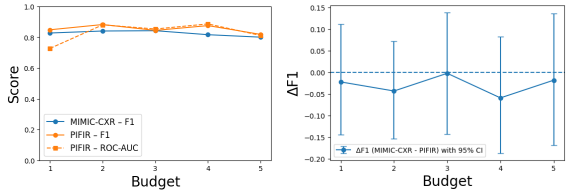

Figure 13: Transfer from MIMIC-CXR to PIFIR under our method RADS.

cells, fluid, bronchial, biopsy, tissue, specimen), reflecting cytology/histopathology reporting that emphasizes sample type and microscopic description rather than imaging observations. PIFIR is characterized by PET-CT and metabolic-imaging terminology (e.g., uptake, FDG, PET, CT, activity), as well as systemic disease descriptors (e.g., marrow, disease), consistent with PET-driven assessment of metabolic activity and whole-body involvement. MIMIC-CXR is dominated by chest radiography vocabulary and common pulmonary findings (e.g., chest, pneumonia, pleural, effusion, pulmonary, lung), reflecting the focus of X-ray reports on thoracic anatomy and acute cardiopulmonary abnormalities. Overall, these TF–IDF profiles highlight substantial modality- and workflow-driven lexical shifts between these datasets, motivating domain-adaptive transfer methods that can operate under pronounced vocabulary mismatch. Figure 3 also shows the transfer gap.

For the CHIFIR and PIFIR datasets, expert annotators also provided span-level annotation of concepts relevant to disease detection. Concept annotations in CHIFIR and PIFIR Datasets are listed in Table 13 and Table 12. The CHIFIR dataset reports 1,155 concepts, and the PIFIR dataset has 3,194 concepts. The two corpora serve different clinical niches. CHIFIR comes from cytology and histopathology notes and therefore focuses on microbiology terms such as `FungalDescriptor` and `Stain`. PIFIR is built from PET–CT reports and centres on imaging findings and risk factors, for example `Abnormality_CT` and `Risk_factor`.

To quantify overlap, we compute the Jaccard Similarity between the concept vocabularies:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{14}$$

where A and B are the sets of surface forms in CHIFIR and PIFIR, respectively. Figure 11 plots the resulting heatmap. Although both datasets include the classification terms `positive`, `equivocal` and `negative`, their lexical realizations share little common ground, so the Jaccard scores remain low.

To quantify directional lexical divergence, we compute the KL divergence on the concept level. For each concept, P and Q denote the distributions of surface forms in PIFIR and CHIFIR, respectively. Both are smoothed over the combined vocabulary $\mathcal{V} = \mathcal{V}_{\text{PIFIR}} \cup \mathcal{V}_{\text{CHIFIR}}$ with a small ε to avoid zeros.

Dataset	Example report excerpt (de-identified)
CHIFIR	*""R groin LN biopsy"". Please note specimen has been received fresh and fragments have been sent to flow cytometry at 1225 on XXXXXX by XX/XXX. Two tan wispy cores 4 and 5 mm in length. A1. (dl:oze) ""R groin LN biopsy"". A tan core, 4mm in length with multiple fragments up to 1mm. A1. (dl/kr)*
PIFIR	*PET/CT technique: Scanning was performed encompassing the base of skull to upper thighs on a PET/CT scanner (GE 690 with time-of-flight). A contemporaneous low dose non-contrast multislice CT scan was performed for anatomic correlation and attenuation correction. Uptake time=70 minutes. BSL=<7mmol/L.*
MIMIC-CXR	*AP upright and lateral views the chest were provided. Mild left basal atelectasis. Lungs are otherwise clear. No signs of pneumonia or edema. No large effusion or pneumothorax. Cardiomediastinal silhouette is normal. Bony structures are intact. No free air below the right hemidiaphragm.*

Table 11: Representative (de-identified) report excerpts from each dataset.

Concept	Count	Unique	Diversity
`Infection_or_IFI`	279	174	0.62
`Risk_factor`	429	179	0.42
`Abnormality`	46	24	0.52
`Abnormality_CT`	460	204	0.44
`Abnormality_PET`	470	224	0.48
`Lung`	372	36	0.10
`Sinus`	19	4	0.21
`Other`	189	92	0.49
`Infection_Inflammation`	354	103	0.29
`IFI_Indication`	37	21	0.57
`improvement`	115	51	0.44
`stable`	29	16	0.55
`worsening`	55	34	0.62
`positive`	124	33	0.27
`equivocal`	129	68	0.53
`negative`	87	23	0.26

Table 12: Summary statistics for the IFI-related concepts in the PIFIR dataset.

Concept	Count	Unique	Diversity
`ClinicalQuery`	68	43	0.63
`FungalDescriptor`	294	86	0.29
`Fungus`	106	19	0.18
`Invasiveness`	39	27	0.69
`Stain`	172	16	0.09
`SampleType`	198	64	0.32
`positive`	118	40	0.34
`equivocal`	8	6	0.75
`negative`	152	12	0.08

Table 13: Summary statistics for the IFI-related concepts in the CHIFIR dataset.

$$KL(P \,\|\, Q) = \sum_{v \in \mathcal{V}} P(v) \, \log \frac{P(v)}{Q(v)}, \qquad (15)$$

where

$$P(v) = \frac{\text{count}_{\text{PIFIR}}(v) + \varepsilon}{\sum_{u \in \mathcal{V}} \text{count}_{\text{PIFIR}}(u) + \varepsilon |\mathcal{V}|}. \qquad (16)$$

We compute $\mathrm{KL}(\text{CHIFIR} \| \text{PIFIR})$ and visualize the results as a heatmap (Figure 9). In KL divergence, larger values indicate that greater mismatch between the two datasets. Even for shared classification terms (`positive`, `equivocal`, `negative`), the divergence values remain large. This suggests that the two datasets differ systematically in how their concepts are expressed, not simply in whether specific words occur.

Representative Report Examples To qualitatively illustrate domain- and modality-specific language, we provide representative (de-identified) report excerpts from each corpus in Table 11.