COHORT-BASED ACTIVE MODALITY ACQUISITION

Anonymous authors
Paper under double-blind review

ABSTRACT

Real-world multimodal machine learning often faces missing, costly-to-acquire modalities, raising the problem of which samples to prioritize for additional acquisition under a budget. Prior work mainly studies per-sample or training-time acquisition while test-time, cohort-level acquisition is less explored. We propose Cohort-based Active Modality Acquisition (CAMA), a novel test-time cohort-level modality acquisition setting, and introduce imputation-based acquisition strategies that estimate the expected utility of acquiring a missing modality, along with upper-bound heuristics for benchmarking. Experiments on datasets with up to 15 modalities demonstrate that our proposed imputation-based strategies can more effectively guide the acquisition of an additional modality for selected samples compared with methods relying solely on pre-acquisition information, entropy-based guidance, or random selection. We showcase the real-world relevance and scalability of our method by demonstrating its ability to guide the acquisition of proteomics data for disease prediction in a large prospective cohort, the UK Biobank (UKB). Our work provides an effective approach for optimizing modality acquisition at the cohort level, enabling more effective use of resources in constrained settings.[1]

1 INTRODUCTION

Consider a clinical healthcare setting where all patients in a cohort undergo a standard, inexpensive set of initial examinations, such as basic blood tests and anamnesis. However, a more advanced, expensive, or invasive procedure, like genomic sequencing or specialized imaging, could offer crucial diagnostic or prognostic information for a subset of these patients (Huang et al., 2021). Given a limited budget or capacity for the more advanced procedure, the central question becomes: which patients should receive this additional resource to maximize the overall diagnostic yield or improve treatment outcomes across the entire cohort?

For healthcare, budgets are often resource-specific rather than flexible. For example, a hospital may have a fixed capacity for one MRI scanner, or a cohort may have a specific grant for one modality. The critical decision is prioritizing access to that single resource across the cohort, and not necessarily dynamically acquiring for different modalities per patient. Concretely, for the UKB, only 10% of the cohort do contain the proteomics modality. Notably, the proteomics modality was financed through a multi-million dollar pre-competitive consortium of 13 pharmaceutical companies, and a $\pounds 20$ million government top-up was later needed to finish the full cohort acquisition (UK Biobank, 2025; Vanderbilt University Medical Center, 2026; UK Biobank, 2023). This highlights the remarkable need of CAMA in real-life applications*, i.e.,* the acquisition of a single, fixed modality for multimodal learning setups. Consider a healthcare system that can afford 1,000 expensive tests for a 100,000-person cohort. The goal is to improve health outcomes across the whole population, and this typically happens through resource allocation: who receives preventive interventions, who gets enrolled in clinical trials, who is flagged for closer monitoring. These decisions depend on accurate risk stratification. A global ranking of all 100,000 individuals by predicted risk becomes the tool through which such allocation decisions are made. The question is: which 1,000 patients should we test so that our final ranking of all 100,000 is as accurate as possible?

Balancing potential gains from data modalities against the costs and complexities of acquisition is not unique to healthcare. In remote sensing, for instance, decisions must be made regarding which geographical areas warrant costly high-resolution satellite imagery to supplement widely

[1] Code will be published on GitHub.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

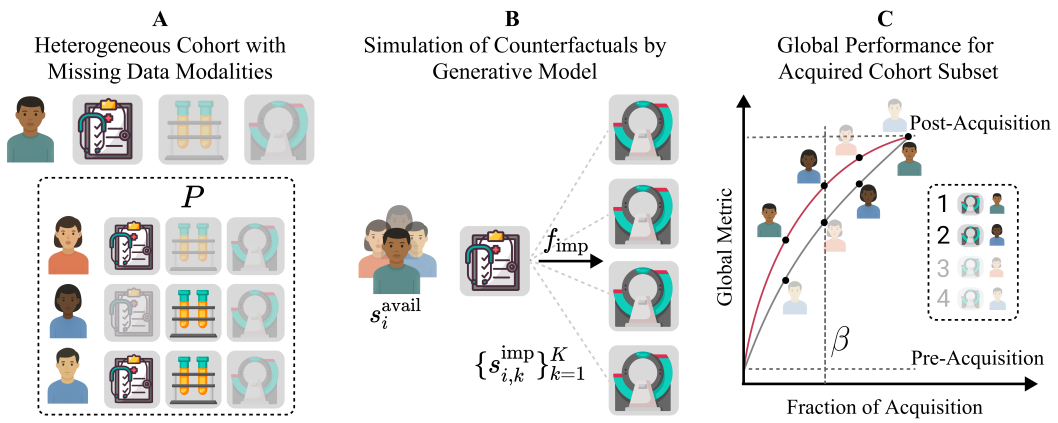


Figure 1: Motivational example for CAMA determining the added value of obtaining the magnetic resonance image (MRI) modality. **(A)** A heterogeneous cohort for which each sample has P distinct modalities. **(B)** Instead of using the initial subset logit scores $\mathrm{s}_i^{\text{avail}}$, a generative model f_{imp} imputes the target missing modality for every patient in the cohort. This yields imputed, augmented-modality logit scores $\{\mathrm{s}_{i,k}^{\text{imp}}\}_{k=1}^{K}$ that approximate the logits as if that modality were available. These scores approximate $\mathrm{s}_i^{\text{acquired}}$, *i.e.,* the counterfactual with only the imputed modality added. **(C)** An acquisition function (AF) utilizes these scores to rank samples by acquisition priority. The graph demonstrates how the global performance metric improves from the initial baseline towards the performance of a model with access to post-acquisition data, as an increasing fraction of the cohort receives the additional modality. This acquisition process is guided by the proposed strategies operating under the acquisition budget constraint β.

available, lower-resolution data. Likewise, in industrial quality assurance, manufacturers could decide which components from a production batch should undergo detailed, time-consuming testing in addition to rapid, standard visual inspections to effectively identify defects at a batch level. The topic of efficient data acquisition has led to several established paradigms in machine learning, such as Active Learning (AL) (Holzmüller et al., 2023), Active Feature Acquisition (AFA) (Shim et al., 2018), Active Modality Acquisition (AMA) (Kossen et al., 2023), and multimodal learning with missing data (Wu et al., 2024). However, previous research predominantly centers on optimizing acquisition for individual samples and often does not directly address test-time budget constraints for an entire cohort. Consequently, the strategic, test-time acquisition of an additional modality from a cohort perspective remains a significant, largely unaddressed gap. This setting involves deciding, for a given batch of new samples where different subsets of modalities are available, which specific samples should receive an additional, costly modality to best achieve a global objective, *e.g.,* maximizing overall predictive performance or diagnostic accuracy for the cohort, subject to budget constraints. We hypothesize that imputation-based acquisition functions (AFs) can effectively guide resource allocation under cohort-level constraints. The main contributions of this work are as follows:

- **The CAMA setting** We introduce and formalize CAMA, a previously unexplored setting that addresses the challenge of prioritizing which samples within a test-time cohort should undergo additional modality acquisition based on an available subset of modalities.
- **Development of AFs for CAMA** We propose a theoretical framework, derived from established evaluation metrics, *e.g.,* Area Under the Receiver Operating Characteristic (AUROC) and Area Under the Precision-Recall Curve (AUPRC), that provides a foundation for developing AFs within the CAMA setting.
- **Architectures for CAMA** We develop a novel architecture for approaching CAMA, including a) derivations of AFs by combining generative and discriminative deep learning and b) the definition of corresponding upper bounds to serve as performance benchmarks.
- **Comprehensive evaluation** We present a comprehensive empirical evaluation of our proposed methods across several multimodal datasets, which vary in their number of modalities and application domains, with up to 100,000 samples and 15 modalities.

2 Related Work

In the following, we contextualize our work on CAMA by reviewing the key concepts and contributions from several relevant research domains summarized briefly in Table 1.

Table 1: Comparison of acquisition paradigms. CAMA is unique due to cohort-level acquisition.

Paradigm	Acquisition	Decision Level	Time	Primary Objective
AL	Labels	Individual	Training	Maximize model performance
AFA	Features	Individual	Test	Optimize sample-level prediction
AMA	Modalities	Individual	Test	Optimize sample-level prediction
CAMA	Modalities	Cohort	Test	Maximize global cohort metric

Active Learning (AL) AL seeks to enhance model training by selecting unlabeled data points for annotation by an oracle (Settles, 2012; Ren et al., 2022; Li et al., 2025). Our methodology draws significantly from AL principles, particularly in the development of an AF to guide the selection process. Consequently, established AL strategies and concepts, such as those rooted in measuring uncertainty (Settles, 2012; Han & Kang, 2021; Hoarau et al., 2025; Raj & Bach, 2022; Ma et al., 2019) or using generative models (Tran et al., 2019; Zhu & Bento, 2017; Zhang et al., 2024; Ma et al., 2019; Peis et al., 2022), are central to our work. Existing work on multimodal acquisition (Rudovic et al., 2019; Das et al., 2022), batch-level selection (Ash et al., 2020; Kirsch et al., 2019; Holzmüller et al., 2023), and balanced AL (Aggarwal et al., 2020; Shen et al., 2023; Zhang et al., 2023; Hoarau et al., 2025) is especially relevant. Our approach diverges from the conventional goals of directly optimizing model training or seeking labels for specific data points: We aim to identify samples for which the acquisition of an additional data modality would be most beneficial given cohort-level constraints.

Active Feature Acquisition (AFA) AFA builds upon AL by selecting the most informative individual features for a given sample, often considering their acquisition costs (Rahbar et al., 2025). Similar to AL approaches, methods for AFA encompass a diverse range of techniques, including strategies based on measuring uncertainty (Hoarau et al., 2025; Astorga et al., 2024), the use of generative models (Li & Oliva, 2021; 2024; Gong et al., 2019; Zannone et al., 2019), and Reinforcement Learning (RL) (Valancius et al., 2024; Janisch et al., 2020; Kleist et al., 2025; Shim et al., 2018; Baja et al., 2025). Other common methodologies involve batch-level perspectives (Asgaonkar et al., 2024), leveraging information bottlenecks (Norcliffe et al., 2025), or employing the Kullback-Leibler Divergence (KL-Divergence) (Natarajan et al., 2018). Some AFA techniques rely on gradient calculations (Ghosh & Lan, 2023), while distinct approaches are formulated as individual, sequential recommender systems (Freyberg et al., 2024; Vivar et al., 2020). At an application level, even Large Language Models (LLMs), such as Med-PaLM 2 (Singhal et al., 2025), could be employed for AFA, although such deployments remain unexplored in this context. While our setting shares the core idea of AFA, it differs significantly: We are not concerned with the selection of individual features, but rather with identifying which entire data modalities to acquire. Furthermore, this decision-making process is applied at the cohort level, rather than optimizing for individual samples.

Active Modality Acquisition (AMA) AMA can be conceptualized as an extension of AFA, distinguished by its focus on selecting entire data modalities rather than individual features or labels. Prominent related research includes approaches employing RL for multimodal data (Kossen et al., 2023; Jain et al., 2025; Li & Oliva, 2025) and methods utilizing submodular optimization in conjunction with Shapley values (Shapley, 1953; He et al., 2024). The approach by Kossen et al. (2023) differs from ours through its reliance on RL, whereas He et al. (2024) primarily investigate how modalities affect optimal learning performance. Further studies have explored the use of Gaussian mixtures within Bayesian optimal experimental design to enhance data acquisition efficiency for model training (Long, 2022). This objective differs from ours, as our focus is not on improving the model training process itself, but rather on optimizing performance for a downstream task at test time. The relative sparsity of existing work for AMA underscores the significance of the research gap that our proposed setting, *i.e.,* CAMA, aims to address.

Multimodal Learning with Missing Data Modalities Research in multimodal learning with missing data modalities offers techniques for robustly handling incomplete datasets. These methods are broadly classified into strategy design aspects*, i.e.,* architecture-focused designs and model combinations, and data processing aspects*, i.e.,* representation learning and modality imputation (Wu et al., 2024). Acknowledging the utility of these approaches, our work emphasizes imputation-based strategies, and thus this paragraph highlights those methods. Imputation of missing features is commonly performed using Auto Encoders (AEs) (Hinton & Zemel, 1993), Variational Auto Encoders (VAEs) (Kingma & Welling, 2014), Generative Adversarial Networks (GANs) (Goodfellow et al., 2014), or Denoising Diffusion Probabilistic Models (DDPMs) (Ho et al., 2020; Rombach et al., 2022). These methods naturally extend to multiple modalities, for example, with VAE-based (Wesego & Rooshenas, 2024; Sutter et al., 2021; Lewis et al., 2021) and DDPM-based (Wang et al., 2023) approaches. Notably, the latter*, i.e., IMDer* (Wang et al., 2023), a multimodal deep learning architecture that imputes missing values with DDPMs in latent spaces, is adapted in our work (Section 5). However, this research area focuses on handling absent modalities rather than deciding which ones to acquire.

3 Problem Formulation

Let $\mathcal{D} = \{(\boldsymbol{x}_i, \mathrm{y}_i)\}_{i=1}^{N}$ be a dataset of N samples. For each sample i, the full feature set $\boldsymbol{x}_i$ is composed of P distinct data modalities, $\boldsymbol{x}_i = \{\boldsymbol{x}_i^{(1)}, \ldots, \boldsymbol{x}_i^{(P)}\}$, and $\mathrm{y}_i \in \{0, 1\}$ is the corresponding binary label. In practice, only a subset of these modalities may be available. We denote the set of indices of available modalities for sample i as $\mathcal{P}_i^{\text{avail}} \subseteq \{1, \ldots, P\}$. Our goal is to decide for which samples to acquire costly missing data to maximize a cohort-level performance metric. This decision is guided by predictive scores (logits), and we consider three key predictive scores for each sample i:

- $\mathrm{s}_i^{\text{avail}}$: The available score by using the subset of data modalities that are observed.
- $\mathrm{s}_i^{\text{acquired}}$: The acquired score by using the available modalities and the acquired modality.
- $\{\mathrm{s}_{i,k}^{\text{imp}}\}_{k=1}^{K}$: A set of K imputed scores that estimate the unknown $\mathrm{s}_i^{\text{acquired}}$ using the available data modalities.

For instance, given the example from Figure 1, in a simple clinical setting with a cheap, universally available base modality*, e.g.,* cardiac biomarkers such as troponin or B-type natriuretic peptide (BNP), and an expensive additional modality*, e.g.,* cardiac MRI, $\mathrm{s}_i^{\text{avail}}$ would be the score from the blood tests alone, while $\mathrm{s}_i^{\text{acquired}}$ would be the score using both tests and MRI. To compute these scores, we assume a single model f parameterized by $\boldsymbol{\theta}$ that can process any subset of modalities. The available and acquired scores are thus:

$$\mathrm{s}_i^{\text{avail}} = f(\boldsymbol{x}_i^{\text{avail}}, \boldsymbol{\theta}) \tag{1}$$

$$\mathrm{s}_i^{\text{acquired}} = f(\boldsymbol{x}_i^{\text{acquired}}, \boldsymbol{\theta}) \tag{2}$$

where $\boldsymbol{x}_i^{\text{avail}}$ and $\boldsymbol{x}_i^{\text{acquired}}$ represent the feature sets for the available and acquired modalities, respectively. To estimate the acquired score without costly acquisition, we use a generative imputation model f_{imp}. This model generates a set of K plausible embeddings that enable the classifier f_C to predict the scores $\{\mathrm{s}_{i,k}^{\text{imp}}\}_{k=1}^{K}$. These imputation-based scores form the basis of our acquisition functions.

The goal of the optimization is to select a subset of samples $\mathcal{S}$ from the cohort of N total samples for which an additional modality should be acquired. This subset $\mathcal{S} \subseteq \{1, \ldots, N\}$ has a predetermined size $|\mathcal{S}| = \beta$, where β is the acquisition budget*, i.e.,* the number of samples for which additional modalities will be acquired. The final score $\mathrm{s}_i(\mathcal{S})$ used for the evaluation of a sample i is then determined by the selection:

$$\mathrm{s}_i(\mathcal{S}) = \begin{cases} \mathrm{s}_i^{\text{acquired}} & \text{if } i \in \mathcal{S} \\ \mathrm{s}_i^{\text{avail}} & \text{if } i \notin \mathcal{S} \end{cases}. \tag{3}$$

The optimization problem is to find the set $\mathcal{S}^*$ that maximizes the chosen performance metric:

$$\mathcal{S}^* = \underset{\mathcal{S} \subseteq \{1,\ldots,N\} : |\mathcal{S}| = \beta}{\arg\max} \text{Metric}(\mathbf{y}, \mathbf{s}(\mathcal{S})) \tag{4}$$

Table 2: Summary of acquisition function (AF) strategies.

Category	Strategies	Inputs	Ranking Criteria
Oracle	AUROC, AUPRC	True labels & acquired scores	Greedy selection for maximum gain.
Upper-Bound	KL-Div., Rank, Uncertainty	True acquired scores	True change in prediction, cohort rank, or uncertainty.
Imputation	KL-Div., Rank, Uncertainty & Probability	Imputed acquired scores	Expected change in prediction, rank, or uncertainty.
Baselines	Uncertainty, Probability	Pre-acquisition scores	Uncertainty or probability using the available modality.
Random	Random	None	Random selection.

where $\mathbf{y} = \{\mathrm{y}_i\}_{i=1}^N$ is the vector of true labels, and $\mathbf{s}(\mathcal{S}) = \{\mathrm{s}_i(\mathcal{S})\}_{i=1}^N$ is the vector of resulting scores for all samples in the cohort. The task is to identify an optimal, constrained subset for which to acquire additional modalities, while maximizing a performance metric across the entire cohort.

4 Acquisition Function Strategies

Directly solving the optimization problem to identify the optimal sample set $\mathcal{S}^*$ is computationally intractable. Therefore, we employ heuristic acquisition functions (AFs) that approximate the optimal selection by ranking samples for modality acquisition. These strategies (Section A) are derived from standard discriminative metrics (Section A.1) and can be categorized as follows (Table 2):

- **Oracle Strategies:** As upper-bound benchmarks, they assume perfect knowledge of outcomes and true labels to greedily select samples yielding the largest gain in the metric.
- **Upper-Bound Heuristic Strategies:** These heuristics assume knowledge of scores under modality completion but are label-agnostic, relying on metrics like the true uncertainty reduction, rank change, or KL-Divergence.
- **Imputation-Based Strategies:** Grounded in counterfactual reasoning, these strategies use a generative model to predict how the score might change if a missing modality were acquired.
- **Baseline Information Strategies:** These strategies only use information from the available modalities*, i.e.,* without imputations, such as its predicted uncertainty or probability.
- **Random Strategy:** A baseline by selecting samples randomly, without any model scores.

Intuitively, these AFs approximate the expected information gain (EIG) from acquiring an additional modality. In our setting, EIG quantifies the expected improvement in a chosen performance metric given the additional information that would become available through a new modality. For evaluation, we introduce a metric that describes the cumulative performance of an AF, normalized by the total possible gain achievable by transitioning all samples to post-acquisition performance (see Figure 1 C, for an illustrative curve). Let $M_{\text{AF}}(b)$ denote the performance curve of an AF strategy for a primary metric M as a function of the budget fraction b of the acquisition budget β, M_{pre} the performance of the pre-acquisition baseline, and M_{post} the performance of the post-acquisition model. The normalized area of gain for an acquisition function AF, which measures the portion of achievable performance gain captured across different budgets, is defined as the area under the performance gain curve, normalized by the maximum possible gain (Equation (5)). Intuitively, a value of 0 indicates no improvement over the pre-acquisition baseline across budgets. A value of 1 indicates matching the post-acquisition performance on average across budgets. Values greater than 1 occur when the cohort's performance at intermediate budgets temporarily exceeds the post-acquisition cohort as detailed later and shown in Figure 3.

$$G^M_{\text{full}}(\text{AF}) = \frac{\int_0^1 (M_{\text{AF}}(b) - M_{\text{pre}})\, db}{M_{\text{post}} - M_{\text{pre}}} \tag{5}$$

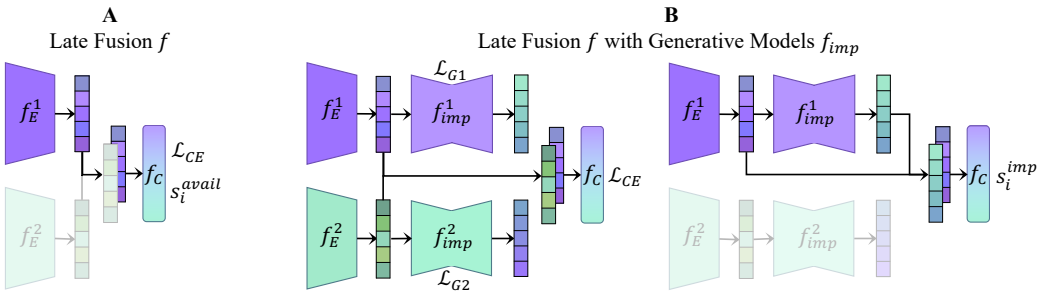


Figure 2: End-to-end architectures to determine the scores for different AFs in our proposed CAMA setting. **(A)** Vanilla late fusion (LF) architecture of a model f that can handle missing data modalities by masking. The model creates scores s_i^{avail} given the available modalities. **(B)** Architecture for training (left) and inference (right) with a late fusion (LF) model f and a generative model f_{imp} to create scores s_i^{imp} for the imputation-based AFs.

5 Experiments

To evaluate the AFs, we require architectures that produce the necessary scores. The oracle, upper-bound, baseline, and random AFs can be evaluated using a vanilla discriminative late fusion (LF) model (Figure 2 A), as they operate on true labels y_i and true scores $\text{s}_i^{\text{acquired}}$ and $\text{s}_i^{\text{avail}}$ (Section 3). In contrast, our proposed imputation-based AFs require the model to predict how its output would change if a missing modality were present. This necessitates an architecture that combines the discriminative component with a generative component to impute the missing modality (Figure 2 B).

Model Architecture First, we implement an architecture with modality-specific encoders $f_E^{(m)} : \mathcal{X}^{(m)} \to \mathbb{R}^d$ and a fusion classifier f_C (Figure 2 A). The encoders map raw inputs for a sample i and modality m to latent embeddings $z_i^{(m)} = f_E^{(m)}(x_i^{(m)})$, *e.g.,* with a Vision Transformer (ViT) (Dosovitskiy et al., 2021) for images and BERT (Devlin et al., 2019) for text. The classifier f_C aggregates these embeddings to produce logits s_i^{avail} depending on the availability of the raw inputs, *i.e.,* with a Transformer encoder (Vaswani et al., 2017). Second, for the imputation-based AFs, we incorporate generative modules additionally to the discriminative late fusion (Figure 2 B) (Wang et al., 2023). The generative modules f_{imp} are parameterized as Diffusion Transformers (DiTs) (Peebles & Xie, 2023) or Beta-Conditional-VAEs (BC-VAEs) (Higgins et al., 2017) trading off performance vs. efficiency. They are trained to approximate the conditional distribution $p_\theta(z^{(k)}|\{z^{(m)}\}_{m\in\mathcal{P}_i^{\text{avail}}})$ for each target modality k. The generative loss for sample i is:

$$\mathcal{L}_{G_i} = \sum_{k\in\mathcal{P}_i^{\text{avail}}} \mathcal{L}_{\text{gen}}\left(z_i^{(k)};\, \{z_i^{(m)}\}_{m\in\mathcal{P}_i^{\text{avail}}}\right) \tag{6}$$

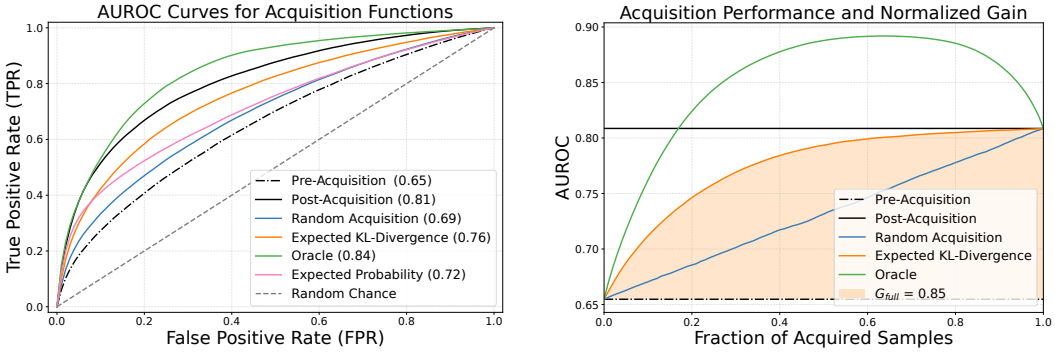
where $\mathcal{L}_{\text{gen}}$ is a variational bound on the negative conditional log-likelihood: for DDPMs, this corresponds to the denoising objective (Ho et al., 2020); for BC-VAEs, this is the negative conditional evidence lower bound (ELBO) (Higgins et al., 2017). For the discriminative task, the loss for sample i is defined as binary cross entropy loss with the label y and the predicted probability p:

$$\mathcal{L}_{\text{CE}_i} = -\left[y_i \log(p_i) + (1-y_i)\log(1-p_i)\right]. \tag{7}$$

The final loss function for the whole architecture is defined as the combination of both loss terms:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{CE}} + \lambda_2 \mathcal{L}_{\text{G}} \tag{8}$$

with loss weightings λ. During inference, the classifier f_C also uses samples of $p(z^{(k)}|\{z^{(m)}\}_{m\in\mathcal{P}_i^{\text{avail}}})$ for missing modalities to create the scores s_i^{imp} needed for our generative AFs (Figure 2 B, right). During training, samples from $p(z^{(k)}|\{z^{(m)}\}_{m\in\mathcal{P}_i^{\text{avail}}})$ are not passed to f_C, even when modalities are missing (Figure 2 B, left). Instead, the discriminative components ($f_E^{(m)}$ and f_C) are trained only on available modalities via attention masks. This means that the generative and discriminative parts



(a) (b)

Figure 3: **(a)** AUROC curves for AFs on the MOSEI dataset at an acquisition budget of 25% of the dataset size. **(b)** Gain achieved during the progressive acquisition of modalities. The oracle AF can exceed the post-acquisition AUROC before declining towards it again.

are trained jointly, but the generative outputs do not directly influence the classifier during training beyond the shared encoders being updated by the classification loss. For model training, we use the ScheduleFree optimizer (Defazio et al., 2024). We find the following architectural decisions essential (Table 7): (a) Layer Normalization (Ba et al., 2016) at the end of each modality's encoder to stabilize the generative models operating between latent spaces, (b) label smoothing (Szegedy et al., 2016) to produce less overconfident and better-calibrated probability distributions, (c) decoupling the generative modules from the classifier during training and (d) class balancing the training dataset (next paragraph). Regarding missing modalities, we do not pre-train on all available data modalities, in contrast to Wang et al. (2023). We use a predefined, seed-dependent missing-modality mask to control data modality leakage during training unlike batch-dependent masks, which eventually reveal all modalities for every sample across numerous epochs. Details in Section B.

Datasets We evaluate CAMA on the UKB (Sudlow et al., 2015), MIMIC Symile (Saporta et al., 2024), MIMIC HAIM (Soenksen et al., 2022a;b), and MOSEI (Zadeh et al., 2018), which cover diverse domains such as healthcare and emotion recognition. Missing modalities are synthetically created, except for the UKB. We design the datasets for binary classification, resulting in ten binary targets for the MIMIC datasets and one binary target for MOSEI and UKB. We employ random oversampling during training for the originally unbalanced HAIM and Symile (Soenksen et al., 2022a; Zadeh et al., 2018). Importantly, during testing we retain the original imbalanced distributions. We highlight UKB as the most challenging dataset to demonstrate that CAMA scales to a broad multimodal range and large-scale cohorts with approximately 100,000 samples and 15 modalities. We focus on acquiring the costly proteomics modality for predicting the onset of systemic lupus erythematosus (SLE), which benefits from the proteomics acquisition (Yang et al., 2025). Details in Section C.

Experimental Setup For datasets with at least three modalities, we apply five-fold cross-validation. Due to initially noisy results for MIMIC HAIM, we increase the number of folds to ten. For each sample in the test set, the initial score s_i^{avail} is established by randomly assigning a subset of available modalities $\mathcal{P}_i^{\text{avail}}$. This procedure is repeated over several runs for robustness. In each run, every sample is stochastically assigned a new subset $\mathcal{P}_i^{\text{avail}}$. Acquisition is simulated by incrementally increasing the budget β. We focus on tasks where the post-acquisition model demonstrates a performance improvement over the pre-acquisition baseline since a simpler pre-acquisition model can outperform a more complex post-acquisition one, potentially due to the introduction of noisy or conflicting signals. In such cases, the final post-acquisition performance falls below the pre-acquisition baseline, resulting in a negative normalized area of gain. To ensure a meaningful evaluation, we exclude any tasks exhibiting this negative gain from the analysis at the split level. For each budget, the top-ranked samples in $\mathcal{S}$ are considered acquired, and their logits are updated from s_i^{avail} to s_i^{acquired}. Reported results are aggregated across all cross-validation splits, combinations of missing and available modalities, and random runs to ensure robustness.

Table 3: Acquisition performance on Symile, with G_{full} shown for AUROC/AUPRC as an example for the class with the best and worst performance and the mean value of all ten classes. Strategies are grouped by category. Best strategy among proposed ones and baselines in bold for each column.

	Acquisitions by AUROC, $G_{\text{full}} \uparrow \pm$ SEM			Acquisitions by AUPRC, $G_{\text{full}} \uparrow \pm$ SEM		
Strategy	**Cardiomegaly**	**Pneumothorax**	**Mean**	**Lung Lesion**	**Pneumothorax**	**Mean**
Upper Bounds (for reference)						
Oracle	2.787 ± 0.139	9.461 ± 1.049	4.580	2.520 ± 0.250	10.623 ± 0.708	4.231
True KL-Div.	0.885 ± 0.011	0.910 ± 0.054	0.883	0.828 ± 0.073	0.827 ± 0.043	0.871
True Rank	0.878 ± 0.019	0.605 ± 0.053	0.811	0.676 ± 0.088	0.483 ± 0.075	0.776
True Uncert.	0.524 ± 0.025	-0.136 ± 0.065	0.481	0.181 ± 0.067	0.293 ± 0.052	0.450
Imputation-based (proposed)						
KL-Divergence	$\mathbf{0.747 \pm 0.039}$	0.773 ± 0.134	**0.833**	$\mathbf{0.896 \pm 0.146}$	0.581 ± 0.084	**0.777**
Probability	0.350 ± 0.053	$\mathbf{0.898 \pm 0.061}$	0.426	0.320 ± 0.104	$\mathbf{0.965 \pm 0.027}$	0.449
Rank	0.378 ± 0.016	0.115 ± 0.082	0.378	0.564 ± 0.086	0.396 ± 0.054	0.407
Uncertainty	0.450 ± 0.041	0.055 ± 0.060	0.440	0.130 ± 0.053	0.513 ± 0.066	0.444
Baselines (no imputation)						
Uncertainty	0.480 ± 0.013	0.536 ± 0.040	0.480	0.215 ± 0.033	0.811 ± 0.041	0.443
Probability	0.431 ± 0.015	0.536 ± 0.040	0.458	0.756 ± 0.136	0.811 ± 0.041	0.550
Random	0.385 ± 0.015	0.327 ± 0.061	0.376	0.503 ± 0.103	0.527 ± 0.053	0.388

Table 4: Acquisition performance on UKB. AFs are grouped by category. Best strategy among proposed ones and baselines in bold.

	AUROC	AUPRC
Strategy	$G_{\text{full}} \pm$ SEM $\uparrow$	$G_{\text{full}} \pm$ SEM $\uparrow$
Upper Bounds (for reference)		
Oracle	1.141 ± 0.051	1.721 ± 0.315
True KL-Div.	0.978 ± 0.007	0.986 ± 0.005
True Rank	0.887 ± 0.022	0.466 ± 0.110
True Uncert.	0.436 ± 0.088	0.507 ± 0.074
Imputation-based (proposed)		
KL-Divergence	$\mathbf{0.641 \pm 0.029}$	0.658 ± 0.045
Probability	0.535 ± 0.026	$\mathbf{0.713 \pm 0.029}$
Rank	0.437 ± 0.028	0.340 ± 0.114
Uncertainty	0.373 ± 0.053	0.332 ± 0.058
Baselines (no imputation)		
Uncertainty	0.365 ± 0.042	0.556 ± 0.073
Probability	0.365 ± 0.042	0.556 ± 0.073
Random	0.528 ± 0.018	0.485 ± 0.052

Table 5: Acquisition performance on MOSEI. AFs are grouped by category. Best strategy among proposed ones and baselines in bold.

	AUROC	AUPRC
Strategy	$G_{\text{full}} \pm$ SEM $\uparrow$	$G_{\text{full}} \pm$ SEM $\uparrow$
Upper Bounds (for reference)		
Oracle	1.478 ± 0.091	1.666 ± 0.161
True KL-Div.	0.882 ± 0.006	0.838 ± 0.006
True Rank	0.849 ± 0.008	0.806 ± 0.010
True Uncert.	0.663 ± 0.006	0.708 ± 0.005
Imputation-based (proposed)		
KL-Divergence	$\mathbf{0.855 \pm 0.034}$	$\mathbf{0.889 \pm 0.052}$
Probability	0.707 ± 0.037	0.846 ± 0.070
Rank	0.432 ± 0.014	0.457 ± 0.019
Uncertainty	0.630 ± 0.015	0.706 ± 0.037
Baselines (no imputation)		
Uncertainty	0.525 ± 0.005	0.540 ± 0.006
Probability	0.433 ± 0.007	0.543 ± 0.009
Random	0.490 ± 0.004	0.525 ± 0.003

Results In the following, we show the effectiveness of CAMA and the proposed model architecture with downstream tasks (Figure 3, Tables 3 to 5), ablations (Table 7) and the general performance added due to modality acquisitions (Table 6). First, as a sanity check and as expected, the AUROC on every dataset is increased after acquisition of the additional modality (Table 6). Second, oracle AFs serve as an upper bound and achieve the highest performance (Tables 3 to 5). Surprisingly, oracle gains can exceed the value of one, as a strategic mix of pre-acquisition and post-acquisition samples can outperform a purely post-acquisition cohort. To benchmark the acquisition logic itself, we use label-agnostic upper-bound heuristics that access acquired scores s_i^{acquired}. AFs based on KL-Divergence and rank change perform well, indicating that prioritizing large predictive shifts or cohort reordering is an effective heuristic in this setting. The imputation-based KL-Divergence strategy consistently outperforms all other non-oracle methods. This AF effectively identifies samples predicted to have the largest shift in their class probability distribution (Figure 3). In contrast, imputation-based AFs relying on rank change, final uncertainty, or final probability are considerably weaker, suggesting that the change in prediction is more effective than estimating the final state. While our primary results for imputation-based AFs use DDPMs, a BC-VAE variant offers faster inference for a trade-off in performance (Table 8 and Section G). Overall, this confirms the robustness and scalability of our framework in a challenging setting. Our results affirm the superiority of the imputation-based KL-Divergence strategy, which achieved substantial and reliable gains over all baselines and heuristics. Details in Sections H to J.

6 DISCUSSION

We introduce CAMA to address modality acquisition under budget constraints. While imputation-based AFs provide an effective solution, we discuss the key implications in the following.

Table 6: Pre- and post-acquisition AUROC (Details in Section F).

Dataset	$M_{\textbf{pre}}$	$M_{\textbf{post}}$
MOSEI	0.6500 ± 0.0633	0.8009 ± 0.0060
UKB	0.5967 ± 0.0745	0.7214 ± 0.0830
Symile	0.5924 ± 0.0310	0.6221 ± 0.0265
HAIM	0.6783 ± 0.0234	0.6865 ± 0.0196

Oracle Behavior Oracles can exceed a model using post-acquisition data for all samples (Figure 3 (b)), *i.e.,* an underlying predictive model can achieve better global performance with a strategic curation of samples. One hypothesis is that additional modalities may introduce variance, redundancy, or conflicting information that imperfect models cannot reconcile but the oracles can circumvent. Surprisingly, the imputation-based KL-Divergence AF can outperform the corresponding upper-bound heuristic (Table 5). Conversely, the performance gap between the rank-change heuristic and its imputation-based counterpart suggests that global, rank-based metrics may be vulnerable to imputation noise. While the KL-Divergence AF demonstrated strong performance, not all imputation-based AFs outperformed simpler AFs (Sections I and J). This indicates that optimal AFs can be context-dependent and that imputation quality is not the only factor.

Imputation Quality and Modality Relationships We impute latent embeddings optimized for the discriminative task rather than raw data (Figure 2). Consequently, standard imputation metrics*, e.g.,* Fréchet inception distance (FID), are not applicable across different generative models. Every generative model influences the encoders latent spaces indirectly. While we observe that utilizing stronger generative models results in higher acquisition performance, generative imputation quality is not the only factor (Table 7). Further, the relationship between imputation quality and the degree of shared information between modalities is important. Intuitively, imputation is expected to be easier when the target modality is more strongly predictable from the observed modalities. We show CAMAs robustness to imputation errors since f_{imp} models a distribution of plausible outcomes rather than aiming for a single reconstruction. By averaging the expected impact across this distribution, the acquisition decision becomes less sensitive to uncertainty.

Table 7: Cross-validated ablation on the Symile dataset.

Ablation (acquisitions by AUROC)	$\textbf{G}_{\textbf{full}} \uparrow$
KL-Divergence (w.r.t. Table 3)	0.833
w/o Layer Norm	0.772
w/o label smoothing	0.746
w/o decoupled data flow	0.599
w/o balanced train set	0.568

Limitations We study acquisition of a single specified modality under a fixed cohort-level budget, which is meaningful for resource-specific settings (Section 1) but does not address the more general problem of jointly selecting both the modality and the sample (Section D). We use randomized missingness patterns as a controlled setup, which facilitates comparison across datasets but does not capture the missingness found in real-world applications. We also do not model heterogeneous acquisition costs across modalities or individuals (Section D). Finally, we do not include a dedicated fairness analysis which may lead to further interesting insights.

7 CONCLUSION AND FUTURE WORK

We introduce CAMA, a novel setting addressing the real-world challenge of optimizing global discriminative performance through strategic test-time acquisition of an additional modality under resource constraints. Our evaluation across multiple multimodal datasets shows that imputation-based AFs can effectively guide resource allocation under cohort-level constraints. In settings such as healthcare, strategic allocation of costly or invasive diagnostic procedures is essential, and our approach offers a promising direction for these applications. Future work includes extending CAMA to multi-class problems or regression tasks, exploring additional imputation techniques, and dynamically selecting which modality to acquire instead of pre-selecting one.

AI USE STATEMENT

In this work, we used generative AI tools to implement methods and to assist with translation. We have not used generative AI tools to generate synthetic data sets, help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, propose or refine hypotheses, design or provide feedback on methodology or experiments, clean or reformat datasets, support qualitative or thematic data analysis, or interpret results. The remaining required disclosure tasks are not applicable to this work. Additionally, we used generative AI tools to create or edit software code and to edit parts of the research paper to improve readability. We did not use generative AI tools for the other activities for which disclosure is recommended. We have reviewed all AI-assisted work. Generative AI was used during the implementation of methods and the creation or editing of software code to support the development and refinement of the required code. All AI-assisted code and implementations were manually reviewed, executed, and checked for correctness and consistency with the intended methodology before being incorporated into the final work. Generative AI was also used to assist with translation. AI-generated or AI-assisted translations were manually reviewed and revised where necessary to ensure that the intended meaning, technical terminology, and context were preserved. Finally, generative AI was used to improve the readability of parts of the manuscript, including aspects such as phrasing, grammar, and clarity. All such edits were reviewed by the authors to ensure that they did not alter the scientific meaning, claims, interpretation, or conclusions of the work. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

The CAMA setting introduced in this paper offers potential for positive broader impacts, primarily by enabling more efficient use of resources in multimodal machine learning. In resource-constrained fields like healthcare, this could facilitate access to more robust and comprehensive model performance by strategically guiding the acquisition of costly or limited additional data modalities. This could translate to improved diagnostic accuracy where such data is critical but not uniformly available for all samples in a cohort. However, the deployment of CAMA, particularly its core function of ranking and prioritizing samples for modality acquisition, necessitates careful ethical consideration. This raises concerns about equity and fairness, especially if the downstream application impacts critical decisions. A significant risk is the potential to introduce biases, including racial, socioeconomic, or other demographic biases. Therefore, the development and application of CAMA must be approached with a strong commitment to ethical principles.

REPRODUCIBILITY STATEMENT

We have taken several steps to facilitate reproducibility of our results. The complete source code used for all experiments will be made publicly available on GitHub upon publication. The repository will include scripts for model training and evaluation. All hyperparameters, including learning rates, batch sizes, and model-specific parameters, are explicitly listed in Section B. Additionally, we provide the complete sweep configurations used for hyperparameter tuning to allow for full replication of our optimization process. Three of the four datasets used in our evaluation are publicly available. For more details see Section 5 and Section C. We provide a full section in Section B and a dedicated paragraph in Section 5 describing implementation details that we found to be crucial.

REFERENCES

Umang Aggarwal, Adrian Popescu, and Céline Hudelot. Active Learning for Imbalanced Datasets. In *IEEE Winter Conference on Applications of Computer Vision, WACV 2020, Snowmass Village, CO, USA, March 1-5, 2020*, pp. 1417–1426. IEEE, 2020. doi: 10.1109/WACV45572.2020.9093475. URL https://doi.org/10.1109/WACV45572.2020.9093475.

Vedang Asgaonkar, Aditya Jain, and Abir De. Generator Assisted Mixture of Experts for Feature Acquisition in Batch. In Michael J. Wooldridge, Jennifer G. Dy, and Sriraam Natarajan (eds.), *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational*

Advances in Artificial Intelligence, EAAI 2014, February 20-27, 2024, Vancouver, Canada, pp. 10927–10934. AAAI Press, 2024. doi: 10.1609/AAAI.V38I10.28967. URL https://doi.org/10.1609/aaai.v38i10.28967.

Jordan T. Ash, Chicheng Zhang, Akshay Krishnamurthy, John Langford, and Alekh Agarwal. Deep Batch Active Learning by Diverse, Uncertain Gradient Lower Bounds. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL https://openreview.net/forum?id=ryghZJBKPS.

Nicolás Astorga, Tennison Liu, Nabeel Seedat, and Mihaela van der Schaar. Active Learning with LLMs for Partially Observed and Cost-Aware Scenarios. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang (eds.), *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024. URL https://papers.nips.cc/paper_files/paper/2024/file/24f3041067ab24157912330344dc3bbd-Paper-Conference.pdf.

Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. Layer Normalization, July 2016. URL http://arxiv.org/abs/1607.06450. arXiv:1607.06450 [stat].

Hilmy Baja, Michiel Kallenberg, and Ioannis N. Athanasiadis. To Measure or Not: A Cost-Sensitive, Selective Measuring Environment for Agricultural Management Decisions with Reinforcement Learning, January 2025. URL http://arxiv.org/abs/2501.12823. arXiv:2501.12823 [cs].

Gyanendra Das, Xavier Thomas, Anant Raj, and Vikram Gupta. MAViC: Multimodal Active Learning for Video Captioning, December 2022. URL http://arxiv.org/abs/2212.11109. arXiv:2212.11109 [cs].

Aaron Defazio, Xingyu Yang, Ahmed Khaled, Konstantin Mishchenko, Harsh Mehta, and Ashok Cutkosky. The Road Less Scheduled. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang (eds.), *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024. URL http://papers.nips.cc/paper_files/paper/2024/hash/136b9a13861308c8948cd308ccd02658-Abstract-Conference.html.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Jill Burstein, Christy Doran, and Thamar Solorio (eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pp. 4171–4186. Association for Computational Linguistics, 2019. doi: 10.18653/V1/N19-1423. URL https://doi.org/10.18653/v1/n19-1423.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *9th International Conference on Learning Representations, ICLR 2021*, 2021.

Jan Freyberg, Abhijit Guha Roy, Terry Spitz, Beverly Freeman, Mike Schaekermann, Patricia Strachan, Eva Schnider, Renee Wong, Dale R. Webster, Alan Karthikesalingam, Yun Liu, Krishnamurthy Dvijotham, and Umesh Telang. MINT: A wrapper to make multi-modal and multi-image AI models interactive, January 2024. URL http://arxiv.org/abs/2401.12032. arXiv:2401.12032 [cs].

Aritra Ghosh and Andrew S. Lan. DiFA: Differentiable Feature Acquisition. In Brian Williams, Yiling Chen, and Jennifer Neville (eds.), *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, pp. 7705–7713. AAAI Press, 2023. doi: 10.1609/AAAI.V37I6.25934. URL https://doi.org/10.1609/aaai.v37i6.25934.

Wenbo Gong, Sebastian Tschiatschek, Sebastian Nowozin, Richard E Turner, José Miguel Hernández-Lobato, and Cheng Zhang. Icebreaker: Element-wise Efficient Information Acquisition with a Bayesian Deep Latent Gaussian Model. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d' Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/c055dcc749c2632fd4dd806301f05ba6-Paper.pdf.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. Generative Adversarial Nets. In Zoubin Ghahramani, Max Welling, Corinna Cortes, Neil D. Lawrence, and Kilian Q. Weinberger (eds.), *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pp. 2672–2680, 2014. URL https://proceedings.neurips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html.

Jongmin Han and Seokho Kang. Active learning with missing values considering imputation uncertainty. *Knowledge-Based Systems*, 224:107079, July 2021. ISSN 09507051. doi: 10.1016/j.knosys.2021.107079. URL https://www.sciencedirect.com/science/article/abs/pii/S0950705121003427?via%3Dihub.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*, pp. 1026–1034. IEEE Computer Society, 2015. doi: 10.1109/ICCV.2015.123. URL https://doi.org/10.1109/ICCV.2015.123.

Yifei He, Runxiang Cheng, Gargi Balasubramaniam, Yao-Hung Hubert Tsai, and Han Zhao. Efficient Modality Selection in Multimodal Learning. *Journal of Machine Learning Research*, 25(47):1–39, 2024. URL http://jmlr.org/papers/v25/23-0439.html.

Irina Higgins, Loïc Matthey, Arka Pal, Christopher P. Burgess, Xavier Glorot, Matthew M. Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. URL https://openreview.net/forum?id=Sy2fzU9gl.

Geoffrey E. Hinton and Richard S. Zemel. Autoencoders, Minimum Description Length and Helmholtz Free Energy. In Jack D. Cowan, Gerald Tesauro, and Joshua Alspector (eds.), *Advances in Neural Information Processing Systems 6, [7th NIPS Conference, Denver, Colorado, USA, 1993]*, pp. 3–10. Morgan Kaufmann, 1993. URL http://papers.nips.cc/paper/798-autoencoders-minimum-description-length-and-helmholtz-free-energy.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (eds.), *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html.

Arthur Hoarau, Benjamin Quost, Sébastien Destercke, and Willem Waegeman. Reducing Aleatoric and Epistemic Uncertainty through Multi-modal Data Acquisition, January 2025. URL http://arxiv.org/abs/2501.18268. arXiv:2501.18268 [cs].

David Holzmüller, Viktor Zaverkin, Johannes Kästner, and Ingo Steinwart. A Framework and Benchmark for Deep Batch Active Learning for Regression. *J. Mach. Learn. Res.*, 24:164:1–164:81, 2023. URL https://jmlr.org/papers/v24/22-0937.html.

Yu Huang, Chenzhuang Du, Zihui Xue, Xuanyao Chen, Hang Zhao, and Longbo Huang. What Makes Multi-Modal Learning Better than Single (Provably). In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp. 10944–10956, 2021. URL

https://proceedings.neurips.cc/paper/2021/hash/5aa3405a3f865c10f420a4a7b55cbff3-Abstract.html.

Eeshaan Jain, Johann Wenckstern, Benedikt von Querfurth, and Charlotte Bunne. Test-Time View Selection for Multi-Modal Decision Making. In *ICLR 2025 Workshop on Machine Learning for Genomics Explorations*, 2025. URL https://openreview.net/forum?id=aNmZ9s6BZV.

Jaromír Janisch, Tomáš Pevný, and Viliam Lisý. Classification with costly features as a sequential decision-making problem. *Machine Learning*, 109(8):1587–1615, August 2020. ISSN 0885-6125, 1573-0565. doi: 10.1007/s10994-020-05874-8. URL http://link.springer.com/10.1007/s10994-020-05874-8.

Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. In Yoshua Bengio and Yann LeCun (eds.), *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014. URL http://arxiv.org/abs/1312.6114.

Andreas Kirsch, Joost van Amersfoort, and Yarin Gal. BatchBALD: Efficient and Diverse Batch Acquisition for Deep Bayesian Active Learning. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché Buc, Emily B. Fox, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pp. 7024–7035, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/95323660ed2124450caaac2c46b5ed90-Abstract.html.

Henrik von Kleist, Alireza Zamanian, Ilya Shpitser, and Narges Ahmidi. Evaluation of Active Feature Acquisition Methods for Time-varying Feature Settings. *Journal of Machine Learning Research*, 26(60):1–84, 2025. URL http://jmlr.org/papers/v26/23-1635.html.

Jannik Kossen, Cătălina Cangea, Eszter Vértes, Andrew Jaegle, Viorica Patraucean, Ira Ktena, Nenad Tomasev, and Danielle Belgrave. Active Acquisition for Multimodal Temporal Data: A Challenging Decision-Making Task. *Transactions on Machine Learning Research*, 2023. URL https://arxiv.org/pdf/2211.05039.

Sarah Lewis, Tatiana Matejovicova, Yingzhen Li, Angus Lamb, Yordan Zaykov, Miltiadis Allamanis, and Cheng Zhang. Accurate Imputation and Efficient Data Acquisition with Transformer-based VAEs. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021*, 2021. URL https://openreview.net/pdf?id=N_OwBEYTcKK.

Dongyuan Li, Zhen Wang, Yankai Chen, Renhe Jiang, Weiping Ding, and Manabu Okumura. A Survey on Deep Active Learning: Recent Advances and New Frontiers. *IEEE Trans. Neural Networks Learn. Syst.*, 36(4):5879–5899, 2025. doi: 10.1109/TNNLS.2024.3396463. URL https://arxiv.org/abs/2405.00334.

Yang Li and Junier Oliva. Active Feature Acquisition with Generative Surrogate Models. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pp. 6450–6459. PMLR, 2021. URL http://proceedings.mlr.press/v139/li21p.html.

Yang Li and Junier Oliva. Distribution Guided Active Feature Acquisition, October 2024. URL http://arxiv.org/abs/2410.03915. arXiv:2410.03915 [cs].

Yang Li and Junier Oliva. Towards Cost Sensitive Decision Making. In Yingzhen Li, Stephan Mandt, Shipra Agrawal, and Mohammad Emtiyaz Khan (eds.), *International Conference on Artificial Intelligence and Statistics, AISTATS 2025, Mai Khao, Thailand, 3-5 May 2025*, volume 258 of *Proceedings of Machine Learning Research*, pp. 3601–3609. PMLR, 2025. URL https://proceedings.mlr.press/v258/li25h.html.

Quan Long. Multimodal information gain in Bayesian design of experiments. *Comput. Stat.*, 37(2): 865–885, 2022. doi: 10.1007/S00180-021-01145-9. URL https://doi.org/10.1007/s00180-021-01145-9.

Chao Ma, Sebastian Tschiatschek, Konstantina Palla, José Miguel Hernández-Lobato, Sebastian Nowozin, and Cheng Zhang. EDDI: Efficient Dynamic Discovery of High-Value Information with Partial VAE. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pp. 4234–4243. PMLR, 2019. URL http://proceedings.mlr.press/v97/ma19c.html.

Sriraam Natarajan, Srijita Das, Nandini Ramanan, Gautam Kunapuli, and Predrag Radivojac. On Whom Should I Perform this Lab Test Next? An Active Feature Elicitation Approach. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, pp. 3498–3505, Stockholm, Sweden, July 2018. International Joint Conferences on Artificial Intelligence Organization. ISBN 978-0-9992411-2-7. doi: 10.24963/ijcai.2018/486. URL https://www.ijcai.org/proceedings/2018/486.

Alexander Luke Ian Norcliffe, Changhee Lee, Fergus Imrie, Mihaela van der Schaar, and Pietro Lio. Information Bottleneck for Active Feature Acquisition, 2025. URL https://openreview.net/forum?id=u5KzZjQOsl.

William Peebles and Saining Xie. Scalable Diffusion Models with Transformers. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*, pp. 4172–4182. IEEE, 2023.

Ignacio Peis, Chao Ma, and José Miguel Hernández-Lobato. Missing Data Imputation and Acquisition with Deep Hierarchical Models and Hamiltonian Monte Carlo. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022. URL http://papers.nips.cc/paper_files/paper/2022/hash/e8dbeb1c947a30576c699e7f5c73d3e3-Abstract-Conference.html.

Arman Rahbar, Linus Aronsson, and Morteza Haghir Chehreghani. A Survey on Active Feature Acquisition Strategies, February 2025. URL http://arxiv.org/abs/2502.11067. arXiv:2502.11067 [cs].

Anant Raj and Francis R. Bach. Convergence of Uncertainty Sampling for Active Learning. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato (eds.), *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pp. 18310–18331. PMLR, 2022. URL https://proceedings.mlr.press/v162/raj22a.html.

Pengzhen Ren, Yun Xiao, Xiaojun Chang, Po-Yao Huang, Zhihui Li, Brij B. Gupta, Xiaojiang Chen, and Xin Wang. A Survey of Deep Active Learning. *ACM Computing Surveys*, 54(9): 1–40, December 2022. ISSN 0360-0300, 1557-7341. doi: 10.1145/3472291. URL https://dl.acm.org/doi/10.1145/3472291.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pp. 10674–10685. IEEE, 2022. doi: 10.1109/CVPR52688.2022.01042. URL https://doi.org/10.1109/CVPR52688.2022.01042.

Ognjen Rudovic, Meiru Zhang, Björn W. Schuller, and Rosalind W. Picard. Multi-modal Active Learning From Human Data: A Deep Reinforcement Learning Approach. In Wen Gao, Helen Mei-Ling Meng, Matthew Turk, Susan R. Fussell, Björn W. Schuller, Yale Song, and Kai Yu (eds.), *International Conference on Multimodal Interaction, ICMI 2019, Suzhou, China, October 14-18, 2019*, pp. 6–15. ACM, 2019. doi: 10.1145/3340555.3353742. URL https://doi.org/10.1145/3340555.3353742.

Adriel Saporta, Aahlad Puli, Mark Goldstein, and Rajesh Ranganath. Contrasting with Symile: Simple Model-Agnostic Representation Learning for Unlimited Modalities. In *Advances in Neural Information Processing Systems*, 2024. URL https://arxiv.org/pdf/2411.01053.

Burr Settles. *Active Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Springer International Publishing, Cham, 2012. ISBN 978-3-031-00432-2 978-3-031-01560-1. doi: 10.1007/978-3-031-01560-1. URL https://link.springer.com/10.1007/978-3-031-01560-1.

L. S. Shapley. A Value for n-Person Games. In Harold William Kuhn and Albert William Tucker (eds.), *Contributions to the Theory of Games (AM-28), Volume II*, pp. 307–318. Princeton University Press, December 1953. ISBN 978-1-4008-8197-0. doi: 10.1515/9781400881970-018. URL https://www.degruyter.com/document/doi/10.1515/9781400881970-018/html.

Meng Shen, Yizheng Huang, Jianxiong Yin, Heqing Zou, Deepu Rajan, and Simon See. Towards Balanced Active Learning for Multimodal Classification. In *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 3434–3445, October 2023. doi: 10.1145/3581783.3612463.

Hajin Shim, Sung Ju Hwang, and Eunho Yang. Joint Active Feature Acquisition and Classification with Variable-Size Set Encoding. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/e5841df2166dd424a57127423d276bbe-Paper.pdf.

Karan Singhal, Tao Tu, Juraj Gottweis, Rory Sayres, Ellery Wulczyn, Mohamed Amin, Le Hou, Kevin Clark, Stephen R. Pfohl, Heather Cole-Lewis, Darlene Neal, Qazi Mamunur Rashid, Mike Schaekermann, Amy Wang, Dev Dash, Jonathan H. Chen, Nigam H. Shah, Sami Lachgar, Philip Andrew Mansfield, Sushant Prakash, Bradley Green, Ewa Dominowska, Blaise Agüera Y Arcas, Nenad Tomašev, Yun Liu, Renee Wong, Christopher Semturs, S. Sara Mahdavi, Joelle K. Barral, Dale R. Webster, Greg S. Corrado, Yossi Matias, Shekoofeh Azizi, Alan Karthikesalingam, and Vivek Natarajan. Toward expert-level medical question answering with large language models. *Nature Medicine*, January 2025. ISSN 1078-8956, 1546-170X. doi: 10.1038/s41591-024-03423-7. URL https://www.nature.com/articles/s41591-024-03423-7.

Luis R. Soenksen, Yu Ma, Cynthia Zeng, Leonard Boussioux, Kimberly Villalobos Carballo, Liangyuan Na, Holly M. Wiberg, Michael L. Li, Ignacio Fuentes, and Dimitris Bertsimas. Integrated multimodal artificial intelligence framework for healthcare applications. *npj Digital Medicine*, 5(1):149, September 2022a. ISSN 2398-6352. doi: 10.1038/s41746-022-00689-4. URL https://www.nature.com/articles/s41746-022-00689-4.

Luis R Soenksen, Yu Ma, Cynthia Zeng, Leonard David Jean Boussioux, Kimberly Villalobos Carballo, Liangyuan Na, Holly Wiberg, Michael Li, Ignacio Fuentes, and Dimitris Bertsimas. Code for generating the HAIM multimodal dataset of MIMIC-IV clinical data and x-rays, 2022b. URL https://physionet.org/content/haim-multimodal/1.0.1/.

Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, Bette Liu, Paul Matthews, Giok Ong, Jill Pell, Alan Silman, Alan Young, Tim Sprosen, Tim Peakman, and Rory Collins. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*, 12(3):e1001779, March 2015. ISSN 1549-1676. doi: 10.1371/journal.pmed.1001779. URL https://dx.plos.org/10.1371/journal.pmed.1001779.

Thomas M. Sutter, Imant Daunhawer, and Julia E. Vogt. Generalized Multimodal ELBO. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=5Y21V0RDBV.

Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the Inception Architecture for Computer Vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

Toan Tran, Thanh-Toan Do, Ian Reid, and Gustavo Carneiro. Bayesian Generative Active Deep Learning. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 6295–6304. PMLR, June 2019. URL https://proceedings.mlr.press/v97/tran19a.html.

UK Biobank. Largest dataset of thousands of proteins marks landmark step for research into human health, October 2023. URL https://www.ukbiobank.ac.uk/news/largest-dataset-of-thousands-of-proteins-marks-landmark-step-for-research-into-human-health/.

UK Biobank. Landmark proteomics project receives £20m from government, June 2025. URL https://www.ukbiobank.ac.uk/news/landmark-proteomics-project-receives-20m-from-government/.

Michael Valancius, Max Lennon, and Junier Oliva. Acquisition Conditioned Oracle for Nongreedy Active Feature Acquisition. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net, 2024. URL https://arxiv.org/pdf/2302.13960.

Vanderbilt University Medical Center. High-throughput biomarker core rates, April 2026. URL https://medsites.vumc.org/highthroughputbiomarkercore/pricing.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pp. 5998–6008, 2017.

Gerome Vivar, Kamilia Mullakaeva, Andreas Zwergal, Nassir Navab, and Seyed-Ahmad Ahmadi. Peri-Diagnostic Decision Support Through Cost-Efficient Feature Acquisition at Test-Time. In Anne L. Martel, Purang Abolmaesumi, Danail Stoyanov, Diana Mateus, Maria A. Zuluaga, S. Kevin Zhou, Daniel Racoceanu, and Leo Joskowicz (eds.), *Medical Image Computing and Computer Assisted Intervention - MICCAI 2020*, volume 12262, pp. 572–581. Springer International Publishing, Cham, 2020. ISBN 978-3-030-59712-2 978-3-030-59713-9. doi: 10.1007/978-3-030-59713-9_55. URL https://link.springer.com/10.1007/978-3-030-59713-9_55. Series Title: Lecture Notes in Computer Science.

Yuanzhi Wang, Yong Li, and Zhen Cui. Incomplete Multimodality-Diffused Emotion Recognition. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023.

Daniel Wesego and Pedram Rooshenas. Score-Based Multimodal Autoencoder. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=JbuP6UV3Fk.

Renjie Wu, Hu Wang, Hsiang-Ting Chen, and Gustavo Carneiro. Deep Multimodal Learning with Missing Modality: A Survey, October 2024. URL http://arxiv.org/abs/2409.07825. arXiv:2409.07825 [cs].

Sisi Yang, Yuanyuan Zhang, Ziliang Ye, Yanjun Zhang, Xiaoqin Gan, Yu Huang, Hao Xiang, Yiting Wu, Yiwei Zhang, and Xianhui Qin. Plasma proteomics for risk prediction and identification of novel drug targets in systemic lupus erythematosus. *Rheumatology*, 64(6):4032–4040, June 2025. ISSN 1462-0324, 1462-0332. doi: 10.1093/rheumatology/keaf055. URL https://academic.oup.com/rheumatology/article/64/6/4032/7990913.

Amir Zadeh, Paul Pu Liang, Soujanya Poria, E. Cambria, and Louis-Philippe Morency. Multimodal Language Analysis in the Wild: CMU-MOSEI Dataset and Interpretable Dynamic Fusion Graph. In *Annual Meeting of the Association for Computational Linguistics*, 2018. URL https://aclanthology.org/P18-1208.pdf.

Sara Zannone, Jose Miguel Hernandez Lobato, Cheng Zhang, and Konstantina Palla. ODIN: Optimal Discovery of High-value INformation Using Model-based Deep Reinforcement Learning. In *Real-world Sequential Decision Making Workshop, ICML*, June 2019. URL https://www.microsoft.com/en-us/research/publication/odin-optimal-discovery-of-high-value-information-using-model-based-deep-reinforcement-learning/.

Jifan Zhang, Shuai Shao, Saurabh Verma, and Robert D. Nowak. Algorithm Selection for Deep Active Learning with Imbalanced Datasets. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. URL http://papers.nips.cc/paper_files/paper/2023/hash/1e77af93008ee6cd248a31723ce357d8-Abstract-Conference.html.

Xulu Zhang, Wengyu Zhang, Xiaoyong Wei, Jinlin Wu, Zhaoxiang Zhang, Zhen Lei, and Qing Li. Generative active learning for image synthesis personalization. In *ACM Multimedia 2024*, 2024.

Jia-Jie Zhu and José Bento. Generative Adversarial Active Learning, November 2017. URL http://arxiv.org/abs/1702.07956. arXiv:1702.07956 [cs].

A Details about Acquisition Function Strategies

A.1 AUROC and AUPRC

To derive the proposed acquisition strategies, we briefly explain the metrics used in the following paragraphs.

AUROC The Area Under the Receiver Operating Characteristic (AUROC) measures the model's ability to discriminate between positive and negative classes and is defined as

$$\text{AUROC}(\mathbf{y}, \mathbf{s}) = \frac{1}{N_+ N_-} \sum_{i:\mathrm{y}_i=1} \sum_{j:\mathrm{y}_j=0} \left(\mathbb{I}(\mathrm{s}_i > \mathrm{s}_j) + \frac{1}{2}\mathbb{I}(\mathrm{s}_i = \mathrm{s}_j) \right) \tag{9}$$

where $N_+ = |\{i \mid \mathrm{y}_i = 1\}|$ and $N_- = |\{j \mid \mathrm{y}_j = 0\}|$.

AUPRC The Area Under the Precision-Recall Curve (AUPRC) summarizes the trade-off between precision (P_t) and recall (R_t) across different decision thresholds t and is defined as

$$\text{AUPRC}(\mathbf{y}, \mathbf{p}) = \sum_{k=1}^{N'} (R_k - R_{k-1}) P_k \tag{10}$$

where points (R_k, P_k) are ordered by threshold from the PR curve, N' is the number of unique thresholds, and $\mathbf{p} = \sigma(\mathbf{s})$.

A.2 Oracle Acquisition Strategies: Exact Gain Calculation

Oracle acquisition strategies serve as theoretical upper limits for the performance of greedy acquisition approaches. They operate under the ideal assumption that the true labels y_i and the outcome scores $\mathrm{s}_i^{\text{acquired}}$ are known for all samples $i \in \{1, \ldots, N\}$. While not implementable in practice, these oracle strategies provide benchmarks by selecting samples based on their exact marginal contribution to the global evaluation metric. The general principle is to iteratively select β samples. At each step, among the samples for which the additional modality has not yet been acquired, the oracle picks the one that provides the largest true immediate gain to the chosen global metric.

AUROC Oracle The AUROC oracle strategy aims to maximize the cohort's AUROC by identifying, at each step, the sample i that yields the largest immediate increase in this metric if its additional modality were acquired (changing its score from $\mathrm{s}_i^{\text{avail}}$ to $\mathrm{s}_i^{\text{acquired}}$), *i.e.,* a greedy selection. This prospective increase is quantified by the marginal gain g_i^{AUROC}. The components of this gain, $g_i^{\text{AUROC}}(\mathrm{y}_i = 1)$ (for positive samples) and $g_i^{\text{AUROC}}(\mathrm{y}_i = 0)$ (for negative samples), reflect the net change in favorable pairwise score comparisons relative to samples of the other class. Recall the definition of AUROC from Equation (9):

$$\text{AUROC}(\mathbf{y}, \mathbf{s}) = \frac{1}{N_+ N_-} \sum_{i:\mathrm{y}_i=1} \sum_{j:\mathrm{y}_j=0} \left(\mathbb{I}(\mathrm{s}_i > \mathrm{s}_j) + \frac{1}{2}\mathbb{I}(\mathrm{s}_i = \mathrm{s}_j) \right).$$

The total marginal gain for sample i, representing the exact change in the cohort's AUROC value, is then, by considering positive and negative samples and neglecting the normalization factor:

$$\begin{aligned} g_i^{\text{AUROC}}(\mathrm{y}_i = 1) = \sum_{j:\mathrm{y}_j=0} \Bigg(& \mathbb{I}(\mathrm{s}_i^{\text{acquired}} > \mathrm{s}_j^{\text{avail}}) - \mathbb{I}(\mathrm{s}_i^{\text{avail}} > \mathrm{s}_j^{\text{avail}}) \\ & + \frac{1}{2}\Big[\mathbb{I}(\mathrm{s}_i^{\text{acquired}} = \mathrm{s}_j^{\text{avail}}) - \mathbb{I}(\mathrm{s}_i^{\text{avail}} = \mathrm{s}_j^{\text{avail}}) \Big] \Bigg) \end{aligned} \tag{11}$$

$$g_i^{\text{AUROC}}(\mathrm{y}_i = 0) = \sum_{j:\mathrm{y}_j=1} \Bigg(\mathbb{I}(\mathrm{s}_j^{\text{avail}} > \mathrm{s}_i^{\text{acquired}}) - \mathbb{I}(\mathrm{s}_j^{\text{avail}} > \mathrm{s}_i^{\text{avail}}) + \frac{1}{2}\Big[\mathbb{I}(\mathrm{s}_j^{\text{avail}} = \mathrm{s}_i^{\text{acquired}}) - \mathbb{I}(\mathrm{s}_j^{\text{avail}} = \mathrm{s}_i^{\text{avail}})\Big]\Bigg) \tag{12}$$

$$g_i^{\text{AUROC}} = \frac{1}{N_+ N_-} \left(g_i^{\text{AUROC}}(\mathrm{y}_i = 1) \cdot \mathbb{I}(\mathrm{y}_i = 1) + g_i^{\text{AUROC}}(\mathrm{y}_i = 0) \cdot \mathbb{I}(\mathrm{y}_i = 0) \right) \tag{13}$$

AUPRC Oracle The AUPRC oracle strategy seeks to maximize the cohort's AUPRC. It operates by identifying, at each step, the sample i which, if its additional modality were acquired (changing its score from $\mathrm{s}_i^{\text{avail}}$ to $\mathrm{s}_i^{\text{acquired}}$), would yield the largest immediate increase in the global AUPRC value*, i.e.,* a greedy selection. This marginal gain, g_i^{AUPRC}, represents the exact change in the cohort's AUPRC. To calculate the marginal gain for a sample i, we compute the change in the cohort's AUPRC. Let $\mathbf{s}^{\text{current}}$ be the vector of scores for the whole cohort. We define a new vector, $\mathbf{s}^{\text{updated}}$, which is identical to $\mathbf{s}^{\text{current}}$ except that for sample i, the score is changed from $\mathrm{s}_i^{\text{avail}}$ to $\mathrm{s}_i^{\text{acquired}}$. The marginal gain is then:

$$g_i^{\text{AUPRC}} = \text{AUPRC}(\mathbf{y}, \mathbf{p}^{\text{updated}}) - \text{AUPRC}(\mathbf{y}, \mathbf{p}^{\text{current}}) \tag{14}$$

where $\mathbf{p}^{\text{current}} = \sigma(\mathbf{s}^{\text{current}})$ and $\mathbf{p}^{\text{updated}} = \sigma(\mathbf{s}^{\text{updated}})$.

A.3 Upper-Bound Heuristic Strategies

The preceding oracle strategies make the assumption of perfect foresight into both the true labels y_i and the exact outcome scores $\mathrm{s}_i^{\text{acquired}}$. We now introduce a distinct class of upper-bound heuristic strategies. These strategies still presume access to the true future scores $\mathrm{s}_i^{\text{acquired}}$ for any sample i if its additional modality were acquired. However, the following upper-bound heuristics are label-agnostic*, i.e.,* the true label y_i of a candidate sample is not used when determining its priority for acquisition. Consequently, the selection principle for these strategies must rely on how the known change from an initial score $\mathrm{s}_i^{\text{avail}}$ to the future score $\mathrm{s}_i^{\text{acquired}}$ is expected to influence the global evaluation metric, without direct reference to the sample's ground-truth label.

Maximum True Uncertainty Reduction The uncertainty reduction strategy prioritizes acquiring the additional modality for samples where doing so is expected to yield the largest decrease in predictive uncertainty. For each sample i, uncertainty is quantified using the binary entropy $\mathcal{H}(p_i)$ of its predicted probability p_i for the positive class, defined as:

$$\mathcal{H}(p_i) = -p_i \log_2 p_i - (1 - p_i) \log_2 (1 - p_i), \tag{15}$$

The acquisition strategy operates with knowledge of the initial probability $p_i^{\text{avail}} = \sigma(\mathrm{s}_i^{\text{avail}})$ derived from the available modalities, and crucially, the true future probability $p_i^{\text{acquired}} = \sigma(\mathrm{s}_i^{\text{acquired}})$ that would be obtained if the additional modality were acquired (where $\mathrm{s}_i^{\text{acquired}}$ is the oracle score). The acquisition score g_i^{UR} for sample i is then the exact reduction in entropy:

$$g_i^{\text{UR}} = \mathcal{H}(p_i^{\text{avail}}) - \mathcal{H}(p_i^{\text{acquired}}). \tag{16}$$

Samples with higher g_i^{UR} values, indicating a greater expected reduction in uncertainty, are prioritized for modality acquisition.

Maximum True Rank Change This rank change strategy prioritizes samples whose relative standing within the cohort, based on predicted probability of belonging to the positive class, would change most significantly if the additional modality were acquired. For each sample i, we consider

its rank $R(p_i)$ when all N samples in the cohort are ordered by their respective probabilities p_i. The acquisition score g_i^{RC} for sample i is defined as the absolute magnitude of this change in rank:

$$g_i^{\text{RC}} = |R(p_i^{\text{acquired}}) - R(p_i^{\text{avail}})|. \tag{17}$$

Samples exhibiting a higher g_i^{RC} are prioritized for modality acquisition, since they are expected to cause the largest shift in the sample's rank-ordered position relative to its peers.

KL-Divergence The KL-Divergence acquisition strategy aims to identify samples for which acquiring the additional modality would lead to the largest change in the predicted probability distribution. Specifically, it quantifies the divergence from the predicted probability distribution based on the true future score, $P_i^{\text{acquired}} \sim \text{Bernoulli}(p_i^{\text{acquired}})$, back to the initial distribution based on baseline data, $P_i^{\text{avail}} \sim \text{Bernoulli}(p_i^{\text{avail}})$. This is measured by the KL-Divergence $D_{\text{KL}}(P_i^{\text{avail}} \| P_i^{\text{acquired}})$ and can be defined as follows for an acquisition function:

$$g_i^{\text{KLD}} = D_{\text{KL}}\left(P_i^{\text{avail}} \middle\| P_i^{\text{acquired}}\right) \tag{18}$$

$$= p_i^{\text{avail}} \log_2 \frac{p_i^{\text{avail}}}{p_i^{\text{acquired}}} + (1 - p_i^{\text{avail}}) \log_2 \frac{1 - p_i^{\text{avail}}}{1 - p_i^{\text{acquired}}} \tag{19}$$

Samples with a higher g_i^{KLD} are prioritized, as this indicates a greater discrepancy between the prediction based on available data and the prediction that would be made with the additional modality.

A.4 Baseline Information Strategies

Shifting from approaches that leverage oracle knowledge of future scores ($\text{s}_i^{\text{acquired}}$), the present section details methods serving as practical, label-agnostic baselines. They make acquisition decisions based exclusively on information derived from the initially available modality ($\text{s}_i^{\text{avail}}$). A random acquisition strategy serves as a fundamental baseline.

Maximum Baseline Uncertainty The Maximum Baseline Uncertainty strategy is a baseline that prioritizes samples for which the prediction based on the initially available modality is most uncertain. The acquisition score for sample i is directly the binary entropy $\mathcal{H}(p_i^{\text{avail}})$, as defined in Equation (15):

$$g_i^{\text{UU}} = \mathcal{H}(p_i^{\text{avail}}). \tag{20}$$

Samples with a higher g_i^{UU}, *i.e.,* p_i^{avail} closer to 0.5, since the entropy $H(p_i^{\text{avail}})$ is symmetric around $p_i^{\text{avail}} = 0.5$, are selected first.

Maximum Baseline Probability This approach prioritizes acquiring the additional modality for samples that the baseline model already predicts as belonging to the positive class with high confidence. The acquisition score g_i^{UP} for sample i is simply its initial probability p_i^{avail} based on the available modality:

$$g_i^{\text{UP}} = p_i^{\text{avail}}, \tag{21}$$

Samples with a higher g_i^{UP} are prioritized for acquisition.

A.5 Imputation-Based Strategies

Having explored strategies that assume perfect knowledge of the true labels y_i and/or future scores $\text{s}_i^{\text{acquired}}$, and simpler baselines relying only on current information $\text{s}_i^{\text{avail}}$, we now introduce methods aiming to bridge the gap by offering a practical and label-agnostic pathway to modality acquisition. They operate by utilizing an imputation model, f_{imp}, to generate a set of K plausible future scores,

denoted $\{\mathrm{s}_{i,k}^{\text{imp}}\}_{k=1}^{K}$, conditioned on the initially available data $\mathrm{s}_i^{\text{avail}}$. The core principle of these strategies is to then derive acquisition scores from statistics of this imputed score distribution, with the goal of emulating the decision-making process, but without requiring true future knowledge at test time.

Maximum Expected Probability The Maximum Expected Probability strategy prioritizes samples which have the highest average probability of belonging to the positive class after modality acquisition. It relies on the set of K imputed future probabilities $\{\mathrm{p}_{i,k}^{\text{imp}}\}_{k=1}^{K}$, where each $\mathrm{p}_{i,k}^{\text{imp}} = \sigma(\mathrm{s}_{i,k}^{\text{imp}})$ is derived from an imputed future score $\mathrm{s}_{i,k}^{\text{imp}}$. The acquisition score g_i^{eP} for sample i is the mean of these imputed probabilities:

$$g_i^{\text{eP}} = \frac{1}{K}\sum_{k=1}^{K} \mathrm{p}_{i,k}^{\text{imp}}. \tag{22}$$

Samples with a higher g_i^{eP} are selected, representing instances where the imputation model, on average, predicts a high likelihood of being positive if the additional modality were acquired.

Maximum Expected Uncertainty Reduction The Maximum Expected Uncertainty Reduction strategy aims to select samples for which the acquisition of the additional modality is anticipated to yield the largest average decrease in predictive uncertainty (Equation (15)). This strategy considers the initial entropy $\mathcal{H}(p_i^{\text{avail}})$, and the distribution of entropies $\{\mathcal{H}(\mathrm{p}_{i,k}^{\text{imp}})\}_{k=1}^{K}$. The acquisition score g_i^{eUR} is the difference between the initial entropy and the mean of the imputed future entropies:

$$g_i^{\text{eUR}} = \mathcal{H}(p_i^{\text{avail}}) - \frac{1}{K}\sum_{k=1}^{K} \mathcal{H}(\mathrm{p}_{i,k}^{\text{imp}}). \tag{23}$$

Samples with higher g_i^{eUR} are prioritized, indicating a greater expected clarification of the prediction upon acquiring the new modality.

Expected Rank Change The Maximum Expected Rank Change strategy prioritizes samples for which the acquisition of the additional modality is anticipated to cause the largest change in their rank, relative to the initial ranking based on p_i^{avail}. It aims to mirror the "Maximum True Rank Change" strategy by using imputed future probabilities. Let $R(p_i^{\text{avail}})$ denote the rank of sample i when all N samples in the cohort are ordered by their initial probabilities p_j^{avail} (for $j = 1, \ldots, N$). For each of the K imputed future probabilities $\mathrm{p}_{i,k}^{\text{imp}}$ for sample i, let $R(\mathrm{p}_{i,k}^{\text{imp}})$ denote the rank of sample i if its probability were $\mathrm{p}_{i,k}^{\text{imp}}$ while all other samples $j \neq i$ retain their initial probabilities p_j^{avail}. The acquisition score g_i^{eRC} is then the mean of the absolute differences between these imputed future ranks and the initial rank:

$$g_i^{\text{eRC}} = \frac{1}{K}\sum_{k=1}^{K} |R(\mathrm{p}_{i,k}^{\text{imp}}) - R(p_i^{\text{avail}})|. \tag{24}$$

Samples with a higher g_i^{eRC} are selected, as they are expected to experience the largest shift in their rank-ordered position relative to other samples in the cohort upon modality acquisition.

Expected KL-Divergence The Expected KL-Divergence strategy selects samples where the initial probability distribution is expected to diverge most significantly from the future probability distributions derived from the K imputed scores. The acquisition score g_i^{eKLD} is the average KL-Divergence $D_{\text{KL}}(P_i^{\text{avail}} \| P_i^{(\text{imp},k)})$ over the K imputations:

$$g_i^{\text{eKLD}} = \frac{1}{K}\sum_{k=1}^{K} D_{\text{KL}}\left(P_i^{\text{avail}} \middle\| P_i^{(\text{imp},k)}\right). \tag{25}$$

A higher g_i^{eKLD} indicates that, on average, the imputed future predictions substantially differ from the initial baseline prediction, suggesting a significant informational update from acquiring the additional modality.

B Hyperparameters, Model Details and Compute Environment

We employ domain-specific encoders to process the respective modalities: for language inputs, we use a pre-trained BERT model (Devlin et al., 2019), for vision, a Vision Transformer (ViT) (Dosovitskiy et al., 2021). Other data types, *e.g.,* temporal sequences, tabular data, or pre-extracted embeddings, are handled by Transformer encoders (Vaswani et al., 2017). We use well-established hyperparameters from the literature for the modality-specific encoders and only optimize the remaining parameters. Notably, our experiments compared three approaches for normalizing the encoder output: No Normalization, Batch Normalization, and Layer Normalization. We found Layer Normalization to be particularly advantageous, as it both stabilized training convergence and significantly enhanced the performance of the DDPMs. We also evaluated the impact of using only the `CLS` token representation from the encoder versus leveraging the full output sequence. This comparison revealed no substantial effect on performance, suggesting the sufficiency of the `CLS` token representation for our task. Layers in the network are initialized using Kaiming initialization (He et al., 2015) if they were not pre-initialized by the specific encoder architecture. We find this particularly important for stabilizing the DDPMs during the early epochs of end-to-end model training.

We perform hyperparameter sweeps for the remaining parts of the designed model in the following ranges:

- Transformer Head
  - Embedding dimension: `[32, 64, 128, 256, 512, 1024]`
  - Feed-Forward network: `[128, 256, 512, 1024, 2048]`
  - Dropout: `[0, 0.1, 0.2]`
  - Number of heads: `[4, 8, 16]`
  - Number of layers: `[2, 4, 6, 8]`
- DDPMs
  - Embedding dimension: analogous to Transformer head
  - Hidden dimension: `[32, 64, 128, 256, 512, 1024]`
  - Dropout: `[0, 0.1, 0.2]`
  - Number of heads: `[4, 8, 16]`
  - Number of layers: `[2, 4, 6, 8]`
  - Number of steps: `[10, 25, 50, 100, 250, 500]`
- ScheduleFree Optimizer
  - Learning rate: `[1e-1, 1e-2, 1e-3, 3e-4, 1e-4, 1e-5]`
  - Warmup steps: `[0, 100, 200]`
  - Weight decay: `[0, 0.01, 0.001]`

The models are trained with early stopping but without any maximum number of epochs. For the imputation-based acquisition functions, 100 DDPM samples are used during inference of the model.

Our experiments are conducted on a High-Performance Cluster (HPC) with the following environment:

1. 21 Dell PowerEdge R7525 compute nodes, each with:
   - 64 AMD Epyc cores (Rome)
   - 512GB RAM
   - 1 NVIDIA A100 40G GPU
2. 2 Dell PowerEdge XE8545 compute nodes, each with:

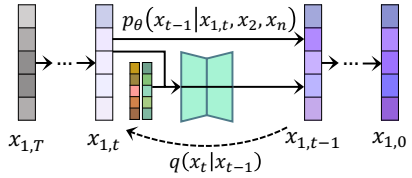


Figure 4: The latent DDPM with its (de)noising functions. Coloring represents less noise in the latent space, starting with pure noise in $X_{i,T} = X_{1,T}$ with T steps. The DDPM is conditioned with two non-missing latent spaces, each from one remaining modality respectively.

- 128 AMD Epyc cores (Milan)
- 512GB RAM
- 4 NVIDIA A100 40G GPUs (NVLink-connected)

C Dataset Details

We evaluate CAMA on four diverse, real-world multimodal datasets, spanning domains from healthcare to emotion recognition.

MIMIC Symile This clinical dataset is derived from the MIMIC database and is designed for predicting the diagnosis of ten classes (Fracture, Enlarged Cardiomediastinum, Consolidation, Atelectasis, Edema, Cardiomegaly, Lung Lesion, Lung Opacity, Pneumonia, Pneumothorax). It contains 10,345 samples from patients in intensive care units. For our experiments, we utilize three distinct modalities: laboratory values, chest X-ray images, and electrocardiograms (ECGs).

MIMIC HAIM This healthcare benchmark also focuses on the diagnostic prediction of ten classes (Fracture, Enlarged Cardiomediastinum, Consolidation, Atelectasis, Edema, Cardiomegaly, Lung Lesion, Lung Opacity, Pneumonia, Pneumothorax). The bimodal dataset consists of 45,050 samples. The two modalities used in our study are laboratory values and chest X-ray images.

CMU-MOSEI This large-scale benchmark targets multimodal sentiment analysis and emotion recognition with seven classes covering different emotions. It contains 22,856 video samples of speakers expressing opinions. The dataset comprises three modalities: vision, acoustics, and language. Notably, unlike the other datasets, we utilize the pre-computed embeddings provided by the authors rather than the raw data.

UK Biobank (UKBB) The UK Biobank is a large-scale, prospective biomedical database from half a million UK participants. In our experiments, the costly modality targeted for acquisition is proteomics, which is available for only a fraction of the full cohort. We constructed a subset of 100,000 samples in which approximately half include proteomics data, accurately simulating a resource-constrained acquisition scenario. The 15 modalities utilized include electronic health records (EHRs), NMR metabolomics, proteomics, physical activity measurements, diet and alcohol consumption questionnaires, baseline characteristics, smoking status, physiological measurements, anthropometry, hand grip strength, cognitive function tests, ECGs, polygenic risk scores (PRS), and arterial stiffness measurements.

D Extension to Flexible Modality Acquisition

In the main paper, we study CAMA in the setting where one specified missing modality is considered for acquisition under a fixed budget β. A natural extension is to allow the acquisition policy to decide *which* missing modality to acquire for each sample, or even to acquire multiple missing modalities per sample. We briefly outline this extension here to clarify that it can be built on top of the present framework, but introduces a substantially more complex optimization problem.

From Fixed-Modality Acquisition to Flexible Modality Selection In the setting considered in the main paper, the optimization problem is to select a subset of samples $\mathcal{S} \subseteq \{1, \dots, N\}$ with $|\mathcal{S}| = \beta$, where the same target modality is acquired for every selected sample. The resulting score for sample i is then

$$\mathrm{s}_i(\mathcal{S}) = \begin{cases} \mathrm{s}_i^{\text{acquired}} & \text{if } i \in \mathcal{S} \\ \mathrm{s}_i^{\text{avail}} & \text{if } i \notin \mathcal{S}. \end{cases}$$

Thus, the acquisition functions only need to rank samples according to the expected utility of acquiring that one target modality. In a flexible multi-modality setting, however, each sample i may have several missing modalities available for acquisition. If $\mathcal{P}_i^{\text{avail}}$ denotes the set of currently available modalities, then the set of candidate missing modalities is

$$\mathcal{P}_i^{\text{miss}} = \{1, \dots, P\} \setminus \mathcal{P}_i^{\text{avail}}.$$

The acquisition decision is no longer only whether sample i should be selected, but also which modality $p \in \mathcal{P}_i^{\text{miss}}$ should be acquired. The decision space therefore changes from ranking samples i to ranking sample-modality pairs (i, p).

Flexible Acquisition with one Modality per Sample A direct extension is to allow acquisition of at most one additional modality per selected sample. In that case, one chooses a set

$$\mathcal{T} \subseteq \{(i, p) : i \in \{1, \dots, N\},\, p \in \mathcal{P}_i^{\text{miss}}\},$$

where each pair (i, p) indicates that modality p is acquired for sample i. If acquisition costs are assumed uniform, one may constrain $|\mathcal{T}| = \beta$. The final score for sample i then depends on whether a pair involving i is selected:

$$\mathrm{s}_i(\mathcal{T}) = \begin{cases} \mathrm{s}_i^{\text{acquired},(p)} & \text{if } (i, p) \in \mathcal{T} \text{ for some } p \\ \mathrm{s}_i^{\text{avail}} & \text{otherwise,} \end{cases}$$

where $\mathrm{s}_i^{\text{acquired},(p)}$ denotes the score obtained when modality p is additionally acquired for sample i. The optimization problem then becomes

$$\mathcal{T}^* = \underset{\mathcal{T}:\, |\mathcal{T}| = \beta}{\arg\max}\, \text{Metric}(\mathbf{y}, \mathbf{s}(\mathcal{T})).$$

Compared to the formulation in the main paper, this extension already increases the search space substantially, since one must now rank all feasible sample-modality pairs rather than only the samples.

Heterogeneous Acquisition Costs Once multiple candidate modalities are allowed, it is natural to consider heterogeneous acquisition costs. Let c_p denote the cost of acquiring modality p. Then the budget constraint becomes

$$\sum_{(i,p) \in \mathcal{T}} c_p \leq \beta,$$

or, more generally, $\sum_{(i,p) \in \mathcal{T}} c_{i,p} \leq \beta$ if acquisition costs depend on both the sample and the modality. In this case, the optimization is no longer a fixed-cardinality subset selection problem, but a knapsack-like combinatorial optimization problem in which the expected performance gain must be traded off against the cost of acquisition.

Acquiring Multiple Modalities per Sample An even more general extension allows several missing modalities to be acquired for the same sample. In this setting, one selects for each sample i a subset

$$\mathcal{A}_i \subseteq \mathcal{P}_i^{\text{miss}},$$

and the resulting score becomes

$$\mathrm{s}_i(\{\mathcal{A}_j\}_{j=1}^N) = f(\boldsymbol{x}_i^{\text{avail}} \cup \{\boldsymbol{x}_i^{(p)} : p \in \mathcal{A}_i\}, \boldsymbol{\theta}).$$

The corresponding optimization problem can be written as

$$\{\mathcal{A}_i^*\}_{i=1}^N = \underset{\{\mathcal{A}_i\}_{i=1}^N}{\arg\max}\, \text{Metric}\big(\mathbf{y}, \mathbf{s}(\{\mathcal{A}_i\}_{i=1}^N)\big)$$

subject to an appropriate budget constraint such as

$$\sum_{i=1}^{N} \sum_{p \in \mathcal{A}_i} c_{i,p} \leq \beta.$$

This is substantially harder than the single-modality setting because the search space grows from subsets of samples to subsets of missing modalities for each sample, which is exponential in the number of candidate modalities.

Application of the Architecture to Extensions Conceptually, the imputation-based framework introduced in the main paper can be extended to this more general setting. For each candidate pair (i, p), one may define an imputed acquired score

$$\{\mathrm{s}_{i,k}^{\mathrm{imp},(p)}\}_{k=1}^{K},$$

obtained by using f_{imp} to generate plausible latent representations for missing modality p, followed by the classifier f_C. These imputed scores can then be used to construct modality-specific acquisition functions

$$\mathrm{AF}(i, p),$$

which estimate the utility of acquiring modality p for sample i. In principle, the same idea could be extended to sets of modalities, although in that case interactions between modalities become important: the utility of acquiring two modalities jointly is generally not equal to the sum of their individual utilities.

Compared to the single-modality formulation studied in the main paper, flexible multi-modality acquisition introduces several additional challenges: (i) the search space expands from ranking samples to ranking sample-modality pairs or subsets of modalities, (ii) heterogeneous acquisition costs must be modeled explicitly, (iii) interactions between acquired modalities may be non-additive, and (iv) evaluating all candidate acquisitions may become computationally prohibitive. We therefore view flexible multi-modality acquisition not as a minor variant of the present problem, but as a substantial extension built on top of CAMA. The single-modality setting studied in the main paper isolates the core cohort-level acquisition problem in a controlled and practically meaningful form, while providing a methodological basis for future work on more general acquisition policies.

E Runtime Analysis

Table 8: Efficiency analysis for different architectures.

Architecture	**Train (sec)** ↓	**Validation (sec)** ↓	**Parameters (M)** ↓
Late fusion	0.02	0.015	86.5
Late fusion w/ DDPMs	0.17	0.16	125
Late fusion w/ BC-VAEs	0.08	0.08	313

F Detailed Model Performance

Table 9: AUROC for detailed task-specific classes in the multi-class datasets.

Dataset	$M_{\textbf{pre}}$	$M_{\textbf{post}}$
Symile - Fracture	0.5907 ± 0.1619	0.6212 ± 0.1842
Symile - Enl. Card.	0.6241 ± 0.1175	0.6667 ± 0.0171
Symile - Consolid.	0.6670 ± 0.0416	0.6862 ± 0.0289
Symile - Atelect.	0.5328 ± 0.0519	0.5897 ± 0.0149
Symile - Edema	0.6868 ± 0.0313	0.6993 ± 0.0103
Symile - Cardiom.	0.6168 ± 0.0443	0.6863 ± 0.0181
Symile - LL	0.4930 ± 0.2566	0.5218 ± 0.1281
Symile - LO	0.5543 ± 0.0685	0.5702 ± 0.0329
Symile - Pleural Eff.	0.5989 ± 0.0230	0.6044 ± 0.0212
Symile - Pneumoth.	0.5594 ± 0.0261	0.5753 ± 0.0227
Haim - Fracture	0.6251 ± 0.1209	0.6244 ± 0.1051
Haim - Enl. Card.	0.6946 ± 0.0347	0.7090 ± 0.0361
Haim - Consolid.	0.7383 ± 0.0228	0.7453 ± 0.0183
Haim - Atelect.	0.6286 ± 0.0397	0.6535 ± 0.0489
Haim - Edema	0.7513 ± 0.0119	0.7675 ± 0.0113
Haim - Cardiom.	0.6963 ± 0.0137	0.7120 ± 0.0144
Haim - LL	0.5933 ± 0.1842	0.5969 ± 0.1829
Haim - LO	0.6329 ± 0.0310	0.6418 ± 0.0390
Haim - Pleural Eff.	0.7082 ± 0.0137	0.7244 ± 0.0145
Haim - Pneumoth.	0.7084 ± 0.0287	0.6840 ± 0.0343

G RESULTS FOR SYMILE WITH BC-VAES

Table 10: Acquisition performance on Symile (AUROC) with Beta-Conditional Variational Auto Encoders. Strategies are grouped by category. Best strategy among proposed ones and baselines in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	2.807 ± 0.326	4.224 ± 0.501	2.716 ± 0.147	5.901 ± 0.861	2.096 ± 0.078	4.423
True KL-Div.	1.009 ± 0.115	0.752 ± 0.095	0.855 ± 0.020	0.689 ± 0.092	0.900 ± 0.007	0.800
True Rank	0.714 ± 0.094	0.728 ± 0.070	0.853 ± 0.017	0.777 ± 0.104	0.890 ± 0.007	0.719
True Uncert.	0.939 ± 0.088	0.735 ± 0.113	0.571 ± 0.024	0.106 ± 0.079	0.719 ± 0.009	0.555
Imputation-based (proposed)						
KL-Div.	0.834 ± 0.060	0.420 ± 0.109	$\mathbf{0.684 \pm 0.017}$	$\mathbf{0.527 \pm 0.077}$	$\mathbf{0.744 \pm 0.013}$	**0.584**
Prob.	0.643 ± 0.073	$\mathbf{0.559 \pm 0.037}$	0.489 ± 0.022	-0.304 ± 0.170	0.603 ± 0.009	0.395
Rank	0.252 ± 0.109	0.281 ± 0.063	0.526 ± 0.019	0.327 ± 0.087	0.444 ± 0.009	0.366
Uncert.	$\mathbf{0.911 \pm 0.073}$	0.557 ± 0.042	0.593 ± 0.023	0.162 ± 0.075	0.637 ± 0.012	0.519
Baselines (no imputation)						
Uncert.	0.862 ± 0.092	0.397 ± 0.052	0.510 ± 0.017	0.423 ± 0.046	0.592 ± 0.008	0.477
Prob.	0.127 ± 0.066	0.508 ± 0.026	0.526 ± 0.016	0.054 ± 0.133	0.462 ± 0.006	0.388
Random	0.429 ± 0.085	0.290 ± 0.060	0.497 ± 0.019	0.102 ± 0.097	0.480 ± 0.006	0.350

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	2.715 ± 0.124	4.677 ± 0.799	6.125 ± 0.911	4.317 ± 0.181	8.654 ± 0.796	4.423
True KL-Div.	0.871 ± 0.011	0.701 ± 0.189	0.582 ± 0.146	0.892 ± 0.014	0.745 ± 0.096	0.800
True Rank	0.843 ± 0.015	0.457 ± 0.225	0.501 ± 0.140	0.825 ± 0.022	0.603 ± 0.102	0.719
True Uncert.	0.701 ± 0.016	0.626 ± 0.130	0.147 ± 0.057	0.664 ± 0.023	0.343 ± 0.036	0.555
Imputation-based (proposed)						
KL-Div.	$\mathbf{0.718 \pm 0.019}$	0.456 ± 0.097	$\mathbf{0.324 \pm 0.232}$	$\mathbf{0.757 \pm 0.025}$	0.380 ± 0.129	**0.584**
Prob.	0.553 ± 0.014	$\mathbf{0.710 \pm 0.213}$	0.023 ± 0.076	0.096 ± 0.030	$\mathbf{0.580 \pm 0.052}$	0.395
Rank	0.416 ± 0.020	0.357 ± 0.309	0.311 ± 0.087	0.493 ± 0.019	0.251 ± 0.065	0.366
Uncert.	0.629 ± 0.018	0.596 ± 0.131	0.095 ± 0.161	0.567 ± 0.019	0.448 ± 0.041	0.519
Baselines (no imputation)						
Uncert.	0.531 ± 0.015	0.534 ± 0.087	-0.022 ± 0.318	0.441 ± 0.017	0.500 ± 0.023	0.477
Prob.	0.424 ± 0.013	0.468 ± 0.137	0.291 ± 0.066	0.518 ± 0.028	0.499 ± 0.023	0.388
Random	0.418 ± 0.014	0.441 ± 0.220	0.152 ± 0.115	0.399 ± 0.018	0.291 ± 0.087	0.350

Table 11: Acquisition performance on Symile (AUPRC) with Beta-Conditional Variational Auto Encoders. Strategies are grouped by category. Best strategy among proposed ones and baselines in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	1.785 ± 0.110	3.468 ± 0.317	2.780 ± 0.146	2.632 ± 0.146	2.460 ± 0.099	4.116
True KL-Div.	0.798 ± 0.050	0.700 ± 0.101	0.828 ± 0.020	0.686 ± 0.026	0.888 ± 0.007	0.766
True Rank	0.736 ± 0.061	0.625 ± 0.096	0.822 ± 0.022	0.733 ± 0.046	0.843 ± 0.009	0.689
True Uncert.	0.778 ± 0.050	0.623 ± 0.044	0.606 ± 0.034	0.206 ± 0.033	0.731 ± 0.011	0.513
Imputation-based (proposed)						
KL-Div.	0.725 ± 0.062	0.574 ± 0.095	$\mathbf{0.636 \pm 0.023}$	$\mathbf{0.599 \pm 0.034}$	$\mathbf{0.733 \pm 0.013}$	**0.624**
Prob.	0.642 ± 0.033	$\mathbf{0.610 \pm 0.037}$	0.539 ± 0.039	0.072 ± 0.059	0.696 ± 0.009	0.433
Rank	0.355 ± 0.050	0.334 ± 0.101	0.409 ± 0.029	0.405 ± 0.032	0.427 ± 0.010	0.409
Uncert.	$\mathbf{0.757 \pm 0.056}$	0.584 ± 0.048	0.592 ± 0.032	0.228 ± 0.030	0.637 ± 0.013	0.480
Baselines (no imputation)						
Uncert.	0.713 ± 0.054	0.239 ± 0.127	0.429 ± 0.024	0.335 ± 0.022	0.579 ± 0.008	0.438
Prob.	0.265 ± 0.039	0.594 ± 0.023	0.567 ± 0.025	0.405 ± 0.035	0.565 ± 0.009	0.523
Random	0.455 ± 0.056	0.248 ± 0.076	0.444 ± 0.027	0.336 ± 0.044	0.485 ± 0.010	0.380

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	2.902 ± 0.184	2.559 ± 0.272	3.156 ± 0.152	5.255 ± 0.253	14.162 ± 1.561	4.116
True KL-Div.	0.871 ± 0.014	0.488 ± 0.088	0.700 ± 0.026	0.879 ± 0.021	0.818 ± 0.044	0.766
True Rank	0.828 ± 0.022	0.533 ± 0.068	0.667 ± 0.036	0.824 ± 0.028	0.278 ± 0.139	0.689
True Uncert.	0.745 ± 0.016	0.183 ± 0.045	0.151 ± 0.032	0.525 ± 0.023	0.579 ± 0.076	0.513
Imputation-based (proposed)						
KL-Div.	$\mathbf{0.749 \pm 0.023}$	0.390 ± 0.117	$\mathbf{0.584 \pm 0.023}$	$\mathbf{0.736 \pm 0.021}$	0.514 ± 0.099	**0.624**
Prob.	0.609 ± 0.016	0.221 ± 0.103	0.013 ± 0.048	0.064 ± 0.037	0.868 ± 0.034	0.433
Rank	0.472 ± 0.015	0.495 ± 0.095	0.336 ± 0.030	0.412 ± 0.026	0.448 ± 0.085	0.409
Uncert.	0.663 ± 0.019	-0.030 ± 0.109	0.202 ± 0.029	0.454 ± 0.021	0.719 ± 0.105	0.480
Baselines (no imputation)						
Uncert.	0.534 ± 0.019	0.021 ± 0.127	0.301 ± 0.025	0.357 ± 0.018	$\mathbf{0.871 \pm 0.076}$	0.438
Prob.	0.478 ± 0.016	$\mathbf{0.549 \pm 0.063}$	0.407 ± 0.037	0.531 ± 0.022	$\mathbf{0.871 \pm 0.076}$	0.523
Random	0.465 ± 0.020	0.320 ± 0.099	0.202 ± 0.030	0.364 ± 0.023	0.483 ± 0.075	0.380

H Detailed Results for MOSEI

Table 12: Acquisition performance on MOSEI (Image imputed by Text), showing G_{full} for AUROC/AUPRC. Strategies are grouped by category. Best strategy among proposed ones and baselines in bold.

	AUROC	**AUPRC**
Strategy	$G_{\text{full}} \pm$ SEM ↑	$G_{\text{full}} \pm$ SEM ↑
Upper Bounds (for reference)		
Oracle	0.995 ± 0.006	0.995 ± 0.012
True KL-Div.	0.777 ± 0.005	0.790 ± 0.015
True Rank	0.763 ± 0.012	0.781 ± 0.014
True Uncert.	0.599 ± 0.015	0.673 ± 0.019
Imputation-based (proposed)		
KL-Div	$\mathbf{0.551 \pm 0.012}$	0.590 ± 0.009
Probability	0.473 ± 0.010	$\mathbf{0.598 \pm 0.006}$
Rank	0.524 ± 0.012	0.560 ± 0.012
Uncertainty	0.500 ± 0.021	0.567 ± 0.009
Baselines (no imputation)		
Uncertainty	0.507 ± 0.015	0.565 ± 0.018
Probability	0.451 ± 0.013	0.578 ± 0.009
Random	0.521 ± 0.006	0.575 ± 0.008

Table 13: Acquisition performance on MOSEI (Image imputed by Audio), showing G_{full} for AUROC/AUPRC. Strategies are grouped by category. Best strategy among proposed ones and baselines in bold.

	AUROC	**AUPRC**
Strategy	$G_{\text{full}} \pm$ SEM ↑	$G_{\text{full}} \pm$ SEM ↑
Upper Bounds (for reference)		
Oracle	1.052 ± 0.007	1.011 ± 0.010
True KL-Div.	0.785 ± 0.009	0.803 ± 0.005
True Rank	0.783 ± 0.010	0.802 ± 0.012
True Uncert.	0.672 ± 0.007	0.742 ± 0.006
Imputation-based (proposed)		
KL-Div	0.566 ± 0.011	0.601 ± 0.009
Probability	0.547 ± 0.002	$\mathbf{0.629 \pm 0.004}$
Rank	$\mathbf{0.576 \pm 0.013}$	0.614 ± 0.007
Uncertainty	0.545 ± 0.009	0.586 ± 0.007
Baselines (no imputation)		
Uncertainty	0.545 ± 0.009	0.601 ± 0.014
Probability	0.526 ± 0.014	0.616 ± 0.008
Random	0.553 ± 0.008	0.603 ± 0.005

Table 14: Acquisition performance on MOSEI (Image imputed by Text and Audio), showing G_{full} for AUROC/AUPRC. Strategies are grouped by category. Best strategy among proposed ones and baselines in bold.

	AUROC	**AUPRC**
Strategy	$G_{\text{full}} \pm$ SEM ↑	$G_{\text{full}} \pm$ SEM ↑
Upper Bounds (for reference)		
Oracle	1.321 ± 0.009	1.315 ± 0.012
True KL-Div.	0.979 ± 0.006	0.862 ± 0.006
True Rank	0.960 ± 0.005	0.811 ± 0.006
True Uncert.	0.716 ± 0.004	0.750 ± 0.004
Imputation-based (proposed)		
KL-Div	$\mathbf{0.603 \pm 0.007}$	0.627 ± 0.009
Probability	0.513 ± 0.002	0.610 ± 0.003
Rank	0.463 ± 0.005	0.491 ± 0.005
Uncertainty	0.499 ± 0.004	0.452 ± 0.004
Baselines (no imputation)		
Uncertainty	0.513 ± 0.004	0.480 ± 0.005
Probability	0.534 ± 0.008	$\mathbf{0.676 \pm 0.007}$
Random	0.489 ± 0.002	0.527 ± 0.002

Table 15: Acquisition performance on MOSEI (Text imputed by Image), showing G_{full} for AUROC/AUPRC. Strategies are grouped by category. Best strategy among proposed ones and baselines in bold.

	AUROC	**AUPRC**
Strategy	$G_{\text{full}} \pm$ SEM ↑	$G_{\text{full}} \pm$ SEM ↑
Upper Bounds (for reference)		
Oracle	1.613 ± 0.269	1.493 ± 0.166
True KL-Div.	0.845 ± 0.018	0.843 ± 0.010
True Rank	0.772 ± 0.024	0.784 ± 0.025
True Uncert.	0.633 ± 0.060	0.697 ± 0.030
Imputation-based (proposed)		
KL-Div	$\mathbf{0.900 \pm 0.023}$	$\mathbf{0.896 \pm 0.022}$
Probability	0.806 ± 0.022	0.861 ± 0.023
Rank	0.309 ± 0.089	0.420 ± 0.065
Uncertainty	0.651 ± 0.049	0.747 ± 0.030
Baselines (no imputation)		
Uncertainty	0.466 ± 0.045	0.537 ± 0.014
Probability	0.418 ± 0.055	0.532 ± 0.045
Random	0.417 ± 0.061	0.489 ± 0.051

Table 16: Acquisition performance on MOSEI (Text imputed by Audio), showing G_{full} for AUROC/AUPRC. Strategies are grouped by category. Best strategy among proposed ones and baselines in bold.

	AUROC	**AUPRC**
Strategy	$G_{\text{full}} \pm$ SEM ↑	$G_{\text{full}} \pm$ SEM ↑
Upper Bounds (for reference)		
Oracle	8.867 ± 1.712	16.066 ± 2.592
True KL-Div.	0.645 ± 0.051	0.387 ± 0.136
True Rank	0.494 ± 0.147	0.015 ± 0.262
True Uncert.	0.913 ± 0.121	0.893 ± 0.141
Imputation-based (proposed)		
KL-Div	2.962 ± 0.925	4.582 ± 1.938
Probability	$\mathbf{3.157 \pm 1.229}$	$\mathbf{6.861 \pm 1.806}$
Rank	-0.413 ± 0.421	-1.134 ± 0.463
Uncertainty	1.592 ± 0.350	3.844 ± 0.806
Baselines (no imputation)		
Uncertainty	0.591 ± 0.147	0.473 ± 0.204
Probability	0.662 ± 0.071	0.971 ± 0.014
Random	0.316 ± 0.059	0.376 ± 0.086

Table 17: Acquisition performance on MOSEI (Text imputed by Image and Audio), showing G_{full} for AUROC/AUPRC. Strategies are grouped by category. Best strategy among proposed ones and baselines in bold.

	AUROC	AUPRC
Strategy	$G_{\text{full}} \pm$ SEM ↑	$G_{\text{full}} \pm$ SEM ↑
Upper Bounds (for reference)		
Oracle	1.207 ± 0.012	1.280 ± 0.013
True KL-Div.	0.851 ± 0.002	0.842 ± 0.004
True Rank	0.836 ± 0.004	0.840 ± 0.005
True Uncert.	0.649 ± 0.009	0.691 ± 0.009
Imputation-based (proposed)		
KL-Div	$\mathbf{0.892 \pm 0.002}$	$\mathbf{0.894 \pm 0.004}$
Probability	0.665 ± 0.004	0.727 ± 0.003
Rank	0.489 ± 0.003	0.520 ± 0.003
Uncertainty	0.662 ± 0.008	0.720 ± 0.007
Baselines (no imputation)		
Uncertainty	0.538 ± 0.009	0.567 ± 0.009
Probability	0.379 ± 0.003	0.462 ± 0.004
Random	0.512 ± 0.002	0.531 ± 0.003

Table 18: Acquisition performance on MOSEI (Audio imputed by Image), showing G_{full} for AUROC/AUPRC. Strategies are grouped by category. Best strategy among proposed ones and baselines in bold.

	AUROC	AUPRC
Strategy	$G_{\text{full}} \pm$ SEM ↑	$G_{\text{full}} \pm$ SEM ↑
Upper Bounds (for reference)		
Oracle	1.238 ± 0.124	1.207 ± 0.093
True KL-Div.	0.826 ± 0.014	0.821 ± 0.019
True Rank	0.752 ± 0.030	0.780 ± 0.025
True Uncert.	0.544 ± 0.034	0.627 ± 0.023
Imputation-based (proposed)		
KL-Div	$\mathbf{0.800 \pm 0.012}$	$\mathbf{0.803 \pm 0.017}$
Probability	0.684 ± 0.020	0.737 ± 0.024
Rank	0.326 ± 0.072	0.445 ± 0.052
Uncertainty	0.552 ± 0.035	0.625 ± 0.023
Baselines (no imputation)		
Uncertainty	0.436 ± 0.043	0.502 ± 0.016
Probability	0.374 ± 0.034	0.510 ± 0.034
Random	0.397 ± 0.052	0.478 ± 0.039

Table 19: Acquisition performance on MOSEI (Audio imputed by Text), showing G_{full} for AUROC/AUPRC. Strategies are grouped by category. Best strategy among proposed ones and baselines in bold.

	AUROC	AUPRC
Strategy	$G_{\text{full}} \pm$ SEM ↑	$G_{\text{full}} \pm$ SEM ↑
Upper Bounds (for reference)		
Oracle	4.955 ± 0.417	6.275 ± 0.948
True KL-Div.	0.815 ± 0.043	0.817 ± 0.057
True Rank	0.563 ± 0.073	0.645 ± 0.098
True Uncert.	0.520 ± 0.026	0.604 ± 0.059
Imputation-based (proposed)		
KL-Div	2.305 ± 0.360	2.335 ± 0.499
Probability	$\mathbf{2.306 \pm 0.259}$	$\mathbf{2.571 \pm 0.483}$
Rank	-0.210 ± 0.123	-0.093 ± 0.135
Uncertainty	1.154 ± 0.141	1.626 ± 0.303
Baselines (no imputation)		
Uncertainty	0.519 ± 0.023	0.604 ± 0.044
Probability	0.416 ± 0.007	0.511 ± 0.018
Random	0.326 ± 0.024	0.439 ± 0.023

Table 20: Acquisition performance on MOSEI (Audio imputed by Image and Text), showing G_{full} for AUROC/AUPRC. Strategies are grouped by category. Best strategy among proposed ones and baselines in bold.

	AUROC	**AUPRC**
Strategy	$G_{\text{full}} \pm$ SEM ↑	$G_{\text{full}} \pm$ SEM ↑
Upper Bounds (for reference)		
Oracle	1.215 ± 0.010	1.275 ± 0.011
True KL-Div.	0.865 ± 0.002	0.850 ± 0.004
True Rank	0.833 ± 0.004	0.837 ± 0.005
True Uncert.	0.645 ± 0.007	0.693 ± 0.006
Imputation-based (proposed)		
KL-Div	$\mathbf{0.857 \pm 0.002}$	$\mathbf{0.846 \pm 0.004}$
Probability	0.667 ± 0.003	0.725 ± 0.003
Rank	0.459 ± 0.004	0.489 ± 0.004
Uncertainty	0.646 ± 0.007	0.694 ± 0.007
Baselines (no imputation)		
Uncertainty	0.536 ± 0.008	0.566 ± 0.008
Probability	0.368 ± 0.005	0.462 ± 0.005
Random	0.503 ± 0.003	0.530 ± 0.003

I DETAILED RESULTS FOR MIMIC SYMILE

Table 21: Acquisition performance on MIMIC Symile for AUROC, showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	2.876 ± 0.368	3.722 ± 0.370	2.856 ± 0.147	4.862 ± 0.356	2.793 ± 0.324	4.580
True KL-Div.	1.029 ± 0.165	1.019 ± 0.051	0.885 ± 0.025	0.841 ± 0.045	0.920 ± 0.011	0.883
True Rank	0.963 ± 0.153	0.924 ± 0.058	0.946 ± 0.025	0.802 ± 0.068	0.887 ± 0.022	0.811
True Uncert.	0.915 ± 0.138	0.763 ± 0.052	0.749 ± 0.031	0.152 ± 0.052	0.648 ± 0.012	0.481
Imputation-based (proposed)						
KL-Div	0.838 ± 0.164	$\mathbf{0.882 \pm 0.064}$	$\mathbf{0.706 \pm 0.021}$	$\mathbf{0.779 \pm 0.114}$	$\mathbf{0.893 \pm 0.080}$	**0.833**
Probability	$\mathbf{0.861 \pm 0.118}$	0.638 ± 0.046	0.610 ± 0.021	0.099 ± 0.107	0.514 ± 0.099	0.426
Rank	0.123 ± 0.155	0.331 ± 0.043	0.434 ± 0.026	0.371 ± 0.083	0.352 ± 0.030	0.378
Uncertainty	0.851 ± 0.138	0.686 ± 0.051	0.701 ± 0.029	0.188 ± 0.094	0.588 ± 0.018	0.440
Baselines (no imputation)						
Uncertainty	0.616 ± 0.100	0.482 ± 0.033	0.543 ± 0.024	0.380 ± 0.031	0.539 ± 0.010	0.480
Probability	0.153 ± 0.086	0.519 ± 0.035	0.464 ± 0.020	0.446 ± 0.068	0.421 ± 0.010	0.458
Random	0.241 ± 0.121	0.399 ± 0.044	0.479 ± 0.020	0.250 ± 0.080	0.425 ± 0.017	0.376

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	2.787 ± 0.139	4.974 ± 0.581	6.657 ± 0.649	4.817 ± 0.430	9.461 ± 1.049	4.580
True KL-Div.	0.885 ± 0.011	0.753 ± 0.179	0.750 ± 0.087	0.837 ± 0.043	0.910 ± 0.054	0.883
True Rank	0.878 ± 0.019	0.411 ± 0.216	0.793 ± 0.056	0.902 ± 0.029	0.605 ± 0.053	0.811
True Uncert.	0.524 ± 0.025	0.251 ± 0.177	0.212 ± 0.054	0.728 ± 0.023	-0.136 ± 0.065	0.481
Imputation-based (proposed)						
KL-Div	$\mathbf{0.747 \pm 0.039}$	$\mathbf{1.266 \pm 0.258}$	$\mathbf{0.683 \pm 0.106}$	$\mathbf{0.761 \pm 0.060}$	0.773 ± 0.134	**0.833**
Probability	0.350 ± 0.053	0.190 ± 0.223	-0.075 ± 0.077	0.172 ± 0.142	$\mathbf{0.898 \pm 0.061}$	0.426
Rank	0.378 ± 0.016	0.607 ± 0.150	0.635 ± 0.080	0.437 ± 0.054	0.115 ± 0.082	0.378
Uncertainty	0.450 ± 0.041	0.022 ± 0.173	0.199 ± 0.057	0.658 ± 0.045	0.055 ± 0.060	0.440
Baselines (no imputation)						
Uncertainty	0.480 ± 0.013	0.201 ± 0.179	0.406 ± 0.041	0.615 ± 0.019	0.536 ± 0.040	0.480
Probability	0.431 ± 0.015	0.778 ± 0.212	0.417 ± 0.065	0.416 ± 0.055	0.536 ± 0.040	0.458
Random	0.385 ± 0.015	0.365 ± 0.225	0.381 ± 0.057	0.505 ± 0.038	0.327 ± 0.061	0.376

Table 22: Acquisition performance on MIMIC Symile for AUPRC, showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	2.579 ± 0.343	2.964 ± 0.201	4.784 ± 1.130	3.093 ± 0.201	3.015 ± 0.141	4.231
True KL-Div.	0.883 ± 0.094	0.970 ± 0.038	0.933 ± 0.057	0.781 ± 0.038	0.939 ± 0.013	0.871
True Rank	0.659 ± 0.176	0.858 ± 0.033	0.903 ± 0.090	0.720 ± 0.064	0.897 ± 0.022	0.776
True Uncert.	0.604 ± 0.073	0.812 ± 0.044	0.709 ± 0.036	0.050 ± 0.039	0.645 ± 0.015	0.450
Imputation-based (proposed)						
KL-Div	$\mathbf{0.770 \pm 0.210}$	$\mathbf{0.899 \pm 0.044}$	$\mathbf{0.895 \pm 0.159}$	$\mathbf{0.684 \pm 0.063}$	$\mathbf{0.832 \pm 0.039}$	**0.777**
Probability	0.631 ± 0.073	0.673 ± 0.036	0.503 ± 0.106	0.028 ± 0.069	0.624 ± 0.049	0.449
Rank	0.461 ± 0.188	0.388 ± 0.034	0.421 ± 0.064	0.318 ± 0.053	0.351 ± 0.022	0.407
Uncertainty	0.663 ± 0.108	0.744 ± 0.046	0.611 ± 0.073	0.118 ± 0.034	0.569 ± 0.017	0.444
Baselines (no imputation)						
Uncertainty	0.428 ± 0.155	0.524 ± 0.031	0.509 ± 0.036	0.246 ± 0.018	0.519 ± 0.011	0.443
Probability	0.242 ± 0.068	0.523 ± 0.026	0.546 ± 0.030	0.512 ± 0.064	0.533 ± 0.013	0.550
Random	0.149 ± 0.103	0.423 ± 0.032	0.429 ± 0.097	0.222 ± 0.094	0.448 ± 0.012	0.388

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	2.746 ± 0.110	2.520 ± 0.250	4.092 ± 0.259	5.895 ± 0.406	10.623 ± 0.708	4.231
True KL-Div.	0.853 ± 0.010	0.828 ± 0.073	0.792 ± 0.037	0.906 ± 0.032	0.827 ± 0.043	0.871
True Rank	0.882 ± 0.018	0.676 ± 0.088	0.771 ± 0.045	0.911 ± 0.030	0.483 ± 0.075	0.776
True Uncert.	0.473 ± 0.029	0.181 ± 0.067	0.140 ± 0.029	0.595 ± 0.024	0.293 ± 0.052	0.450
Imputation-based (proposed)						
KL-Div	$\mathbf{0.729 \pm 0.034}$	$\mathbf{0.896 \pm 0.146}$	$\mathbf{0.722 \pm 0.043}$	$\mathbf{0.763 \pm 0.059}$	0.581 ± 0.084	**0.777**
Probability	0.366 ± 0.047	0.320 ± 0.104	0.041 ± 0.065	0.339 ± 0.079	$\mathbf{0.965 \pm 0.027}$	0.449
Rank	0.384 ± 0.014	0.564 ± 0.086	0.402 ± 0.050	0.389 ± 0.034	0.396 ± 0.054	0.407
Uncertainty	0.424 ± 0.038	0.130 ± 0.053	0.149 ± 0.034	0.519 ± 0.033	0.513 ± 0.066	0.444
Baselines (no imputation)						
Uncertainty	0.434 ± 0.018	0.215 ± 0.033	0.260 ± 0.023	0.482 ± 0.019	0.811 ± 0.041	0.443
Probability	0.479 ± 0.017	0.756 ± 0.136	0.555 ± 0.037	0.541 ± 0.029	0.811 ± 0.041	0.550
Random	0.386 ± 0.013	0.503 ± 0.103	0.354 ± 0.053	0.435 ± 0.033	0.527 ± 0.053	0.388

Table 23: Acquisition performance on MIMIC Symile for AUROC (Image imputed by Lab), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	4.144 ± 2.172	2.120 ± 0.465	2.803 ± 0.306	2.381 ± 0.215	1.684 ± 0.070	4.104
True KL-Div.	0.728 ± 0.137	0.816 ± 0.011	0.611 ± 0.084	0.837 ± 0.028	0.813 ± 0.009	0.825
True Rank	1.062 ± 0.643	0.614 ± 0.111	0.591 ± 0.157	0.816 ± 0.065	0.765 ± 0.024	0.707
True Uncert.	0.686 ± 0.374	0.606 ± 0.068	0.637 ± 0.140	0.496 ± 0.148	0.619 ± 0.024	0.601
Imputation-based (proposed)						
KL-Div	-1.410 ± 0.037	0.477 ± 0.108	0.282 ± 0.146	$\mathbf{0.626 \pm 0.137}$	0.503 ± 0.022	0.247
Probability	$\mathbf{0.231 \pm 0.288}$	0.498 ± 0.036	0.514 ± 0.060	0.503 ± 0.056	0.454 ± 0.020	0.379
Rank	-0.725 ± 0.858	0.190 ± 0.201	0.520 ± 0.018	0.403 ± 0.085	$\mathbf{0.527 \pm 0.032}$	0.225
Uncertainty	0.168 ± 0.317	0.400 ± 0.025	$\mathbf{0.620 \pm 0.179}$	0.422 ± 0.130	0.493 ± 0.019	**0.448**
Baselines (no imputation)						
Uncertainty	0.129 ± 0.339	0.332 ± 0.069	0.423 ± 0.201	0.456 ± 0.106	0.503 ± 0.020	0.433
Probability	0.129 ± 0.339	$\mathbf{0.515 \pm 0.057}$	0.530 ± 0.049	0.601 ± 0.101	0.477 ± 0.018	0.444
Random	-0.509 ± 0.041	0.356 ± 0.070	0.503 ± 0.163	0.609 ± 0.091	0.492 ± 0.023	0.307

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	1.793 ± 0.077	5.417 ± 4.001	7.073 ± 2.202	4.743 ± 1.268	8.881 ± 1.894	4.104
True KL-Div.	0.822 ± 0.010	1.543 ± 0.689	0.663 ± 0.095	0.884 ± 0.063	0.530 ± 0.109	0.825
True Rank	0.663 ± 0.042	1.291 ± 0.557	0.530 ± 0.154	0.705 ± 0.170	0.035 ± 0.068	0.707
True Uncert.	0.272 ± 0.091	1.128 ± 0.870	0.622 ± 0.187	0.687 ± 0.038	0.254 ± 0.213	0.601
Imputation-based (proposed)						
KL-Div	$\mathbf{0.492 \pm 0.030}$	$\mathbf{0.847 \pm 0.033}$	0.459 ± 0.146	$\mathbf{0.481 \pm 0.158}$	-0.285 ± 0.221	0.247
Probability	0.301 ± 0.081	0.418 ± 0.193	0.034 ± 0.303	0.372 ± 0.256	0.461 ± 0.128	0.379
Rank	0.434 ± 0.045	-0.219 ± 0.837	$\mathbf{0.956 \pm 0.383}$	0.363 ± 0.080	-0.201 ± 0.261	0.225
Uncertainty	0.395 ± 0.020	0.574 ± 0.327	0.400 ± 0.126	0.395 ± 0.040	$\mathbf{0.608 \pm 0.201}$	**0.448**
Baselines (no imputation)						
Uncertainty	0.433 ± 0.015	0.490 ± 0.096	0.558 ± 0.358	0.424 ± 0.053	0.587 ± 0.180	0.433
Probability	0.419 ± 0.065	0.505 ± 0.108	0.263 ± 0.254	0.417 ± 0.228	0.587 ± 0.180	0.444
Random	0.420 ± 0.063	0.384 ± 0.068	0.200 ± 0.455	0.444 ± 0.085	0.169 ± 0.359	0.307

Table 24: Acquisition performance on MIMIC Symile for AUPRC (Image imputed by Lab), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	3.844 ± 2.595	1.792 ± 0.288	2.826 ± 0.677	1.930 ± 0.221	1.778 ± 0.101	4.862
True KL-Div.	0.869 ± 0.023	0.827 ± 0.021	0.620 ± 0.156	0.812 ± 0.024	0.775 ± 0.013	0.770
True Rank	1.398 ± 0.874	0.566 ± 0.140	0.492 ± 0.345	0.827 ± 0.034	0.759 ± 0.024	0.575
True Uncert.	0.712 ± 0.210	0.650 ± 0.111	0.655 ± 0.118	0.538 ± 0.165	0.579 ± 0.041	0.498
Imputation-based (proposed)						
KL-Div	-1.450 ± 1.121	0.464 ± 0.132	0.149 ± 0.370	$\mathbf{0.652 \pm 0.149}$	0.610 ± 0.010	0.131
Probability	$\mathbf{0.210 \pm 0.451}$	0.575 ± 0.018	0.735 ± 0.084	0.591 ± 0.035	0.598 ± 0.016	0.568
Rank	-0.777 ± 1.263	0.261 ± 0.185	0.548 ± 0.051	0.438 ± 0.134	0.480 ± 0.042	0.290
Uncertainty	0.157 ± 0.477	0.456 ± 0.075	0.516 ± 0.245	0.477 ± 0.170	0.358 ± 0.021	0.441
Baselines (no imputation)						
Uncertainty	0.132 ± 0.484	0.369 ± 0.094	0.199 ± 0.387	0.466 ± 0.152	0.368 ± 0.022	0.401
Probability	0.132 ± 0.484	$\mathbf{0.592 \pm 0.029}$	$\mathbf{0.754 \pm 0.084}$	0.632 ± 0.071	$\mathbf{0.612 \pm 0.012}$	**0.621**
Random	-0.110 ± 0.296	0.408 ± 0.060	0.420 ± 0.208	0.599 ± 0.156	0.489 ± 0.037	0.440

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	1.749 ± 0.092	2.456 ± 1.140	7.505 ± 3.891	4.470 ± 1.249	20.270 ± 6.707	4.862
True KL-Div.	0.774 ± 0.017	1.035 ± 0.214	0.413 ± 0.176	0.793 ± 0.046	0.781 ± 0.044	0.770
True Rank	0.667 ± 0.037	0.888 ± 0.274	0.326 ± 0.248	0.669 ± 0.058	-0.837 ± 0.865	0.575
True Uncert.	0.269 ± 0.134	0.598 ± 0.430	0.291 ± 0.099	0.562 ± 0.027	0.129 ± 0.407	0.498
Imputation-based (proposed)						
KL-Div	$\mathbf{0.507 \pm 0.059}$	$\mathbf{0.742 \pm 0.101}$	0.382 ± 0.251	0.639 ± 0.066	-1.388 ± 1.198	0.131
Probability	0.327 ± 0.126	0.398 ± 0.274	0.407 ± 0.140	0.659 ± 0.129	1.176 ± 0.314	0.568
Rank	0.390 ± 0.068	0.069 ± 0.552	$\mathbf{1.071 \pm 0.611}$	0.469 ± 0.059	-0.047 ± 0.555	0.290
Uncertainty	0.321 ± 0.042	0.317 ± 0.186	0.176 ± 0.196	0.242 ± 0.030	$\mathbf{1.390 \pm 0.419}$	0.441
Baselines (no imputation)						
Uncertainty	0.343 ± 0.033	0.372 ± 0.120	0.142 ± 0.320	0.265 ± 0.044	1.352 ± 0.410	0.401
Probability	0.452 ± 0.103	0.528 ± 0.184	0.497 ± 0.287	$\mathbf{0.664 \pm 0.107}$	1.352 ± 0.410	**0.621**
Random	0.369 ± 0.101	0.478 ± 0.117	0.810 ± 0.961	0.449 ± 0.051	0.484 ± 0.166	0.440

Table 25: Acquisition performance on MIMIC Symile for AUROC (Image imputed by ECG), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	4.028 ± 2.600	1.959 ± 0.358	1.960 ± 0.164	3.526 ± 0.947	1.639 ± 0.050	3.892
True KL-Div.	0.963 ± 0.224	0.852 ± 0.051	0.725 ± 0.030	0.614 ± 0.152	0.777 ± 0.003	0.874
True Rank	0.192 ± 0.075	0.559 ± 0.101	0.753 ± 0.013	-0.784 ± 1.091	0.719 ± 0.043	0.327
True Uncert.	0.934 ± 0.189	0.700 ± 0.131	0.717 ± 0.039	-0.853 ± 0.584	0.658 ± 0.050	0.024
Imputation-based (proposed)						
KL-Div	0.425 ± 0.323	$\mathbf{0.561 \pm 0.079}$	$\mathbf{0.769 \pm 0.096}$	-0.353 ± 0.827	$\mathbf{0.621 \pm 0.030}$	**0.669**
Probability	0.650 ± 0.024	0.538 ± 0.107	0.709 ± 0.095	0.162 ± 0.130	0.591 ± 0.024	0.443
Rank	0.411 ± 0.193	0.337 ± 0.122	0.538 ± 0.050	-0.165 ± 0.412	0.445 ± 0.038	0.380
Uncertainty	$\mathbf{0.750 \pm 0.011}$	0.527 ± 0.112	0.720 ± 0.083	0.466 ± 0.229	0.431 ± 0.026	0.167
Baselines (no imputation)						
Uncertainty	-0.962 ± 1.403	0.380 ± 0.050	0.567 ± 0.064	$\mathbf{0.591 \pm 0.180}$	0.427 ± 0.022	-0.055
Probability	0.133 ± 0.291	0.442 ± 0.050	0.500 ± 0.028	-0.684 ± 0.967	0.427 ± 0.022	0.650
Random	-0.803 ± 0.529	0.482 ± 0.038	0.647 ± 0.019	0.227 ± 0.307	0.513 ± 0.018	0.263

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	2.263 ± 0.151	9.495 ± 7.441	4.112 ± 0.961	2.196 ± 0.132	7.745 ± 2.386	3.892
True KL-Div.	0.777 ± 0.016	1.758 ± 1.206	0.881 ± 0.085	0.725 ± 0.019	0.664 ± 0.088	0.874
True Rank	0.560 ± 0.050	-0.731 ± 0.850	0.808 ± 0.103	0.777 ± 0.022	0.416 ± 0.209	0.327
True Uncert.	0.741 ± 0.021	-3.450 ± 3.932	-0.150 ± 0.160	0.727 ± 0.010	0.213 ± 0.412	0.024
Imputation-based (proposed)						
KL-Div	0.271 ± 0.065	2.123 ± 1.791	$\mathbf{0.776 \pm 0.125}$	0.571 ± 0.018	0.924 ± 0.134	**0.669**
Probability	0.374 ± 0.034	-0.235 ± 0.921	0.122 ± 0.178	0.568 ± 0.030	$\mathbf{0.952 \pm 0.095}$	0.443
Rank	0.201 ± 0.072	0.764 ± 0.047	0.272 ± 0.153	0.639 ± 0.029	0.363 ± 0.307	0.380
Uncertainty	$\mathbf{0.382 \pm 0.028}$	-2.196 ± 2.882	0.086 ± 0.165	0.630 ± 0.021	-0.129 ± 0.190	0.167
Baselines (no imputation)						
Uncertainty	0.307 ± 0.038	-3.361 ± 3.790	0.231 ± 0.070	$\mathbf{0.649 \pm 0.017}$	0.625 ± 0.086	-0.055
Probability	0.336 ± 0.018	$\mathbf{3.511 \pm 2.928}$	0.632 ± 0.027	0.575 ± 0.048	0.625 ± 0.086	0.650
Random	0.222 ± 0.044	0.102 ± 0.610	0.485 ± 0.131	0.515 ± 0.036	0.240 ± 0.153	0.263

Table 26: Acquisition performance on MIMIC Symile for AUPRC (Image imputed by ECG), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	6.343 ± 5.165	1.593 ± 0.113	1.950 ± 0.162	2.785 ± 0.530	1.832 ± 0.041	4.085
True KL-Div.	1.358 ± 0.553	0.818 ± 0.060	0.776 ± 0.039	0.526 ± 0.206	0.855 ± 0.004	0.814
True Rank	-0.157 ± 0.576	0.643 ± 0.064	0.808 ± 0.008	-0.477 ± 0.824	0.741 ± 0.031	0.420
True Uncert.	1.331 ± 0.520	0.699 ± 0.124	0.790 ± 0.026	-0.640 ± 0.325	0.745 ± 0.040	0.489
Imputation-based (proposed)						
KL-Div	-0.784 ± 1.571	$\mathbf{0.613 \pm 0.082}$	0.870 ± 0.127	-0.203 ± 0.570	$\mathbf{0.737 \pm 0.035}$	0.434
Probability	0.498 ± 0.210	0.552 ± 0.123	0.868 ± 0.103	0.136 ± 0.118	0.722 ± 0.027	**0.590**
Rank	0.116 ± 0.576	0.468 ± 0.033	0.652 ± 0.045	-0.068 ± 0.285	0.524 ± 0.027	0.419
Uncertainty	$\mathbf{0.642 \pm 0.147}$	0.553 ± 0.129	$\mathbf{0.874 \pm 0.099}$	0.297 ± 0.183	0.553 ± 0.015	0.462
Baselines (no imputation)						
Uncertainty	-4.523 ± 5.002	0.451 ± 0.068	0.668 ± 0.070	$\mathbf{0.360 \pm 0.105}$	0.534 ± 0.013	-0.049
Probability	-0.190 ± 0.747	0.496 ± 0.049	0.578 ± 0.027	-0.626 ± 0.854	0.534 ± 0.013	0.382
Random	-3.009 ± 2.890	0.530 ± 0.060	0.715 ± 0.037	-0.038 ± 0.461	0.591 ± 0.012	0.098

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	2.458 ± 0.268	2.158 ± 0.482	2.735 ± 0.464	3.162 ± 0.383	15.830 ± 5.601	4.085
True KL-Div.	0.827 ± 0.009	0.679 ± 0.250	0.711 ± 0.063	0.727 ± 0.026	0.861 ± 0.024	0.814
True Rank	0.555 ± 0.041	-0.005 ± 0.389	0.715 ± 0.109	0.809 ± 0.034	0.573 ± 0.150	0.420
True Uncert.	0.803 ± 0.012	0.184 ± 0.270	-0.069 ± 0.111	0.723 ± 0.046	0.326 ± 0.207	0.489
Imputation-based (proposed)						
KL-Div	0.288 ± 0.055	0.468 ± 0.419	0.466 ± 0.061	0.595 ± 0.049	$\mathbf{1.285 \pm 0.172}$	0.434
Probability	0.471 ± 0.038	0.519 ± 0.046	0.173 ± 0.148	$\mathbf{0.712 \pm 0.048}$	1.249 ± 0.094	**0.590**
Rank	0.234 ± 0.057	0.554 ± 0.118	0.214 ± 0.099	0.688 ± 0.053	0.810 ± 0.388	0.419
Uncertainty	$\mathbf{0.477 \pm 0.035}$	0.224 ± 0.340	0.181 ± 0.070	0.625 ± 0.030	0.192 ± 0.082	0.462
Baselines (no imputation)						
Uncertainty	0.304 ± 0.125	0.033 ± 0.364	0.225 ± 0.012	0.659 ± 0.046	0.801 ± 0.041	-0.049
Probability	0.427 ± 0.037	$\mathbf{0.704 \pm 0.458}$	$\mathbf{0.524 \pm 0.057}$	0.574 ± 0.072	0.801 ± 0.041	0.382
Random	0.227 ± 0.042	0.518 ± 0.037	0.362 ± 0.020	0.505 ± 0.048	0.584 ± 0.155	0.098

Table 27: Acquisition performance on MIMIC Symile for AUROC (Image imputed by Lab and ECG), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	2.679 ± 0.496	3.081 ± 0.257	2.539 ± 0.150	5.098 ± 0.624	2.119 ± 0.086	3.911
True KL-Div.	1.006 ± 0.111	1.105 ± 0.082	0.880 ± 0.033	0.703 ± 0.095	1.017 ± 0.023	0.922
True Rank	0.760 ± 0.142	1.076 ± 0.073	0.993 ± 0.052	0.930 ± 0.136	1.036 ± 0.023	0.884
True Uncert.	0.975 ± 0.140	0.877 ± 0.083	0.886 ± 0.065	0.300 ± 0.043	0.634 ± 0.018	0.494
Imputation-based (proposed)						
KL-Div	0.770 ± 0.122	$\mathbf{0.736 \pm 0.077}$	0.711 ± 0.031	0.049 ± 0.191	0.535 ± 0.015	0.512
Probability	$\mathbf{0.883 \pm 0.111}$	0.563 ± 0.062	0.620 ± 0.034	$\mathbf{0.706 \pm 0.096}$	$\mathbf{0.593 \pm 0.010}$	**0.628**
Rank	0.594 ± 0.082	0.335 ± 0.079	0.423 ± 0.070	0.130 ± 0.226	0.466 ± 0.012	0.392
Uncertainty	0.785 ± 0.085	0.535 ± 0.064	$\mathbf{0.728 \pm 0.050}$	0.545 ± 0.088	0.556 ± 0.015	0.505
Baselines (no imputation)						
Uncertainty	0.514 ± 0.151	0.495 ± 0.038	0.529 ± 0.045	0.466 ± 0.056	0.545 ± 0.014	0.474
Probability	0.265 ± 0.046	0.463 ± 0.035	0.461 ± 0.037	0.268 ± 0.136	0.401 ± 0.009	0.421
Random	0.005 ± 0.204	0.386 ± 0.047	0.467 ± 0.031	0.373 ± 0.145	0.472 ± 0.014	0.425

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	2.552 ± 0.116	3.249 ± 0.605	4.964 ± 0.431	3.736 ± 0.369	9.096 ± 1.223	3.911
True KL-Div.	0.910 ± 0.030	0.517 ± 0.243	0.811 ± 0.052	0.967 ± 0.061	1.298 ± 0.121	0.922
True Rank	1.047 ± 0.032	0.545 ± 0.093	0.839 ± 0.055	1.010 ± 0.056	0.607 ± 0.109	0.884
True Uncert.	0.514 ± 0.027	0.200 ± 0.151	0.293 ± 0.058	0.729 ± 0.044	-0.468 ± 0.155	0.494
Imputation-based (proposed)						
KL-Div	0.374 ± 0.016	0.403 ± 0.210	0.457 ± 0.079	0.518 ± 0.036	0.565 ± 0.103	0.512
Probability	0.453 ± 0.016	0.680 ± 0.352	0.391 ± 0.064	0.443 ± 0.042	$\mathbf{0.950 \pm 0.096}$	**0.628**
Rank	0.408 ± 0.017	0.581 ± 0.211	0.493 ± 0.063	0.487 ± 0.054	-0.000 ± 0.092	0.392
Uncertainty	$\mathbf{0.476 \pm 0.011}$	0.289 ± 0.162	0.366 ± 0.052	$\mathbf{0.599 \pm 0.031}$	0.171 ± 0.078	0.505
Baselines (no imputation)						
Uncertainty	0.457 ± 0.013	0.241 ± 0.261	0.381 ± 0.040	0.588 ± 0.023	0.520 ± 0.055	0.474
Probability	0.437 ± 0.011	0.448 ± 0.230	0.496 ± 0.057	0.449 ± 0.027	0.519 ± 0.055	0.421
Random	0.418 ± 0.015	$\mathbf{0.739 \pm 0.256}$	$\mathbf{0.530 \pm 0.100}$	0.493 ± 0.051	0.362 ± 0.101	0.425

Table 28: Acquisition performance on MIMIC Symile for AUPRC (Image imputed by Lab and ECG), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	3.244 ± 0.899	2.660 ± 0.365	3.290 ± 0.320	3.756 ± 0.538	2.484 ± 0.159	4.063
True KL-Div.	1.111 ± 0.223	0.983 ± 1.014	0.998 ± 0.079	0.674 ± 0.088	1.014 ± 0.034	0.939
True Rank	0.503 ± 0.538	0.727 ± 0.084	1.063 ± 0.108	0.762 ± 0.088	1.014 ± 0.036	0.822
True Uncert.	0.826 ± 0.166	0.950 ± 0.103	0.923 ± 0.060	0.143 ± 0.041	0.636 ± 0.029	0.493
Imputation-based (proposed)						
KL-Div	$\mathbf{1.238 \pm 0.634}$	$\mathbf{0.636 \pm 0.075}$	0.754 ± 0.075	0.330 ± 0.146	0.467 ± 0.020	0.585
Probability	0.821 ± 0.184	0.581 ± 0.048	0.714 ± 0.055	$\mathbf{0.478 \pm 0.083}$	$\mathbf{0.648 \pm 0.012}$	**0.610**
Rank	1.133 ± 0.543	0.375 ± 0.081	0.582 ± 0.103	0.139 ± 0.174	0.469 ± 0.018	0.491
Uncertainty	1.018 ± 0.316	0.516 ± 0.058	$\mathbf{0.783 \pm 0.081}$	0.285 ± 0.039	0.505 ± 0.015	0.500
Baselines (no imputation)						
Uncertainty	0.843 ± 0.289	0.410 ± 0.050	0.562 ± 0.061	0.275 ± 0.038	0.521 ± 0.019	0.467
Probability	0.196 ± 0.143	0.567 ± 0.044	0.520 ± 0.042	0.442 ± 0.111	0.572 ± 0.014	0.515
Random	-0.073 ± 0.199	0.397 ± 0.084	0.556 ± 0.053	0.411 ± 0.100	0.470 ± 0.013	0.406

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	2.494 ± 0.181	2.376 ± 0.472	4.061 ± 0.416	4.964 ± 0.520	11.303 ± 0.974	4.063
True KL-Div.	0.823 ± 0.022	0.809 ± 0.139	0.873 ± 0.081	1.002 ± 0.049	1.105 ± 0.058	0.939
True Rank	1.012 ± 0.039	0.573 ± 0.099	0.952 ± 0.093	1.027 ± 0.058	0.591 ± 0.077	0.822
True Uncert.	0.394 ± 0.037	0.201 ± 0.029	0.184 ± 0.050	0.624 ± 0.044	0.054 ± 0.058	0.493
Imputation-based (proposed)						
KL-Div	0.346 ± 0.023	0.560 ± 0.075	0.524 ± 0.074	0.441 ± 0.037	0.551 ± 0.075	0.585
Probability	0.448 ± 0.029	0.569 ± 0.185	0.404 ± 0.073	0.481 ± 0.039	$\mathbf{0.959 \pm 0.048}$	**0.610**
Rank	0.390 ± 0.023	0.523 ± 0.091	0.359 ± 0.088	0.449 ± 0.034	0.492 ± 0.045	0.491
Uncertainty	0.346 ± 0.019	0.323 ± 0.061	0.279 ± 0.041	0.466 ± 0.034	0.475 ± 0.062	0.500
Baselines (no imputation)						
Uncertainty	0.334 ± 0.018	0.257 ± 0.049	0.262 ± 0.035	0.491 ± 0.029	0.715 ± 0.033	0.467
Probability	$\mathbf{0.555 \pm 0.015}$	0.448 ± 0.096	$\mathbf{0.607 \pm 0.062}$	$\mathbf{0.528 \pm 0.032}$	0.715 ± 0.033	0.515
Random	0.415 ± 0.015	$\mathbf{0.575 \pm 0.228}$	0.368 ± 0.069	0.432 ± 0.057	0.510 ± 0.090	0.406

Table 29: Acquisition performance on MIMIC Symile for AUROC (Lab imputed by Image), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	2.404 ± 0.448	5.403 ± 1.547	2.446 ± 0.350	4.261	4.715 ± 0.988	6.306
True KL-Div.	0.901 ± 0.104	0.727 ± 0.113	0.874 ± 0.045	0.807	0.670 ± 0.093	0.634
True Rank	0.490 ± 0.208	0.872 ± 0.019	0.873 ± 0.081	1.691	0.627 ± 0.148	0.604
True Uncert.	-0.204 ± 0.122	0.690 ± 0.193	0.319 ± 0.084	-0.511	0.507 ± 0.100	0.154
Imputation-based (proposed)						
KL-Div	0.443 ± 0.028	0.503 ± 0.316	0.523 ± 0.078	**0.383**	0.193 ± 0.090	**0.684**
Probability	0.087 ± 0.217	0.428 ± 0.155	$\mathbf{0.574 \pm 0.117}$	0.221	0.457 ± 0.056	0.380
Rank	$\mathbf{0.848 \pm 0.131}$	-0.103 ± 0.100	0.309 ± 0.073	-0.390	0.463 ± 0.125	-0.139
Uncertainty	0.074 ± 0.187	0.593 ± 0.168	0.415 ± 0.107	0.278	0.624 ± 0.029	0.311
Baselines (no imputation)						
Uncertainty	0.515 ± 0.062	$\mathbf{0.781 \pm 0.101}$	0.394 ± 0.106	0.310	$\mathbf{0.636 \pm 0.041}$	0.487
Probability	0.105 ± 0.251	0.444 ± 0.144	0.567 ± 0.112	0.267	0.449 ± 0.055	0.390
Random	0.366 ± 0.519	0.662 ± 0.085	0.504 ± 0.111	0.087	0.229 ± 0.142	0.419

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	11.461	9.175 ± 4.619	10.440 ± 4.341	4.058 ± 0.233	8.693 ± 2.295	6.306
True KL-Div.	0.359	0.043 ± 1.152	0.608 ± 0.141	0.793 ± 0.054	0.561 ± 0.157	0.634
True Rank	0.396	0.640 ± 1.426	-0.637 ± 0.091	0.797 ± 0.031	0.296 ± 0.208	0.604
True Uncert.	0.507	1.296 ± 0.137	-1.510 ± 0.311	0.320 ± 0.119	0.128 ± 0.257	0.154
Imputation-based (proposed)						
KL-Div	0.477	$\mathbf{3.593 \pm 0.801}$	0.055 ± 1.005	$\mathbf{0.630 \pm 0.024}$	0.037 ± 0.291	**0.684**
Probability	0.551	2.347 ± 0.438	-2.017 ± 0.592	0.581 ± 0.030	$\mathbf{0.575 \pm 0.116}$	0.380
Rank	-0.413	-1.673 ± 4.164	-0.741 ± 0.142	0.398 ± 0.150	-0.084 ± 0.300	-0.139
Uncertainty	**0.646**	0.200 ± 0.174	-0.684 ± 0.697	0.401 ± 0.073	0.566 ± 0.142	0.311
Baselines (no imputation)						
Uncertainty	0.277	0.172 ± 0.139	$\mathbf{0.706 \pm 0.448}$	0.501 ± 0.107	0.575 ± 0.126	0.487
Probability	0.471	1.854 ± 1.407	-1.445 ± 0.962	0.615 ± 0.016	0.575 ± 0.126	0.390
Random	-0.872	2.030 ± 1.179	0.329 ± 0.460	0.490 ± 0.124	0.365 ± 0.137	0.419

Table 30: Acquisition performance on MIMIC Symile for AUPRC (Lab imputed by Image), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	2.142 ± 0.360	3.772 ± 0.624	4.042 ± 1.747	–	5.560 ± 1.093	4.751
True KL-Div.	0.916 ± 0.056	0.648 ± 0.079	0.997 ± 0.160	–	0.568 ± 0.132	0.699
True Rank	0.400 ± 0.206	0.686 ± 0.095	1.106 ± 0.342	–	0.603 ± 0.139	0.590
True Uncert.	-1.106 ± 0.446	0.506 ± 0.100	0.258 ± 0.136	–	0.477 ± 0.270	0.153
Imputation-based (proposed)						
KL-Div	0.499 ± 0.117	0.481 ± 0.203	0.413 ± 0.187	–	0.269 ± 0.189	**0.680**
Probability	-0.381 ± 0.441	0.618 ± 0.008	0.339 ± 0.318	–	$\mathbf{0.681 \pm 0.114}$	0.442
Rank	$\mathbf{0.824 \pm 0.103}$	-0.157 ± 0.006	-0.044 ± 0.404	–	0.246 ± 0.115	0.186
Uncertainty	-0.429 ± 0.473	0.415 ± 0.161	$\mathbf{0.541 \pm 0.220}$	–	0.570 ± 0.164	0.124
Baselines (no imputation)						
Uncertainty	0.455 ± 0.060	0.484 ± 0.099	0.537 ± 0.231	–	0.585 ± 0.160	0.440
Probability	-0.362 ± 0.445	0.629 ± 0.007	0.325 ± 0.326	–	0.672 ± 0.112	0.572
Random	0.283 ± 0.513	$\mathbf{0.635 \pm 0.001}$	0.303 ± 0.231	–	0.253 ± 0.115	0.356

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	–	3.615 ± 0.753	3.868	4.819 ± 0.587	10.192 ± 4.607	4.751
True KL-Div.	–	0.503 ± 0.458	0.513	0.865 ± 0.112	0.585 ± 0.087	0.699
True Rank	–	0.974 ± 0.661	-0.184	0.709 ± 0.066	0.426 ± 0.076	0.590
True Uncert.	–	0.621 ± 0.157	-0.406	0.202 ± 0.069	0.674 ± 0.151	0.153
Imputation-based (proposed)						
KL-Div	–	$\mathbf{1.905 \pm 0.140}$	**0.987**	$\mathbf{0.723 \pm 0.115}$	0.162 ± 0.123	**0.680**
Probability	–	1.133 ± 0.483	-0.486	0.659 ± 0.050	$\mathbf{0.976 \pm 0.085}$	0.442
Rank	–	0.274 ± 1.456	-0.263	0.304 ± 0.238	0.301 ± 0.113	0.186
Uncertainty	–	-0.397 ± 0.271	-0.893	0.261 ± 0.024	0.925 ± 0.045	0.124
Baselines (no imputation)						
Uncertainty	–	0.053 ± 0.070	0.109	0.330 ± 0.055	0.963 ± 0.071	0.440
Probability	–	1.363 ± 0.660	0.282	0.708 ± 0.061	0.963 ± 0.071	0.572
Random	–	0.088 ± 0.319	0.534	0.455 ± 0.187	0.296 ± 0.199	0.356

Table 31: Acquisition performance on MIMIC Symile for AUROC (Lab imputed by ECG), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	2.830 ± 1.006	6.507 ± 2.813	1.993 ± 0.252	13.751 ± 3.349	5.706 ± 1.964	7.637
True KL-Div.	0.807 ± 0.072	1.043 ± 0.176	0.855 ± 0.056	0.933 ± 0.345	0.841 ± 0.028	0.727
True Rank	0.454 ± 0.258	0.880 ± 0.255	0.756 ± 0.027	0.462 ± 0.641	0.398 ± 0.151	0.290
True Uncert.	0.665 ± 0.104	0.273 ± 0.305	0.688 ± 0.056	-0.218 ± 1.149	0.568 ± 0.041	0.437
Imputation-based (proposed)						
KL-Div	$\mathbf{0.735 \pm 0.130}$	$\mathbf{1.221 \pm 0.208}$	0.670 ± 0.136	$\mathbf{4.739 \pm 1.683}$	2.234 ± 0.684	**2.096**
Probability	0.569 ± 0.115	0.126 ± 0.370	$\mathbf{0.775 \pm 0.075}$	-4.498 ± 1.913	$\mathbf{2.255 \pm 0.646}$	-0.333
Rank	0.395 ± 0.117	0.160 ± 0.439	0.298 ± 0.039	1.303 ± 1.170	-1.119 ± 0.726	0.619
Uncertainty	0.612 ± 0.125	0.499 ± 0.346	0.773 ± 0.078	-4.102 ± 1.614	0.071 ± 0.175	-0.744
Baselines (no imputation)						
Uncertainty	0.421 ± 0.290	0.544 ± 0.303	0.456 ± 0.081	0.431 ± 0.170	0.324 ± 0.024	0.574
Probability	0.069 ± 0.066	0.372 ± 0.420	0.505 ± 0.102	0.161 ± 0.787	0.324 ± 0.024	0.521
Random	0.117 ± 0.418	0.737 ± 0.338	0.531 ± 0.041	-0.400 ± 1.001	0.107 ± 0.142	0.189

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	–	13.135 ± 7.362	9.638 ± 2.483	3.075 ± 0.250	12.098 ± 2.848	7.637
True KL-Div.	–	-0.560 ± 1.556	0.982 ± 0.050	0.761 ± 0.031	0.884 ± 0.076	0.727
True Rank	–	-2.847 ± 4.746	1.545 ± 0.340	0.776 ± 0.020	0.185 ± 0.511	0.290
True Uncert.	–	2.300 ± 1.585	0.157 ± 0.299	0.563 ± 0.120	-1.060 ± 0.540	0.437
Imputation-based (proposed)						
KL-Div	–	$\mathbf{4.580 \pm 1.600}$	$\mathbf{2.539 \pm 0.511}$	$\mathbf{0.646 \pm 0.117}$	1.494 ± 0.303	**2.096**
Probability	–	-2.903 ± 3.257	-1.391 ± 0.383	0.455 ± 0.124	$\mathbf{1.617 \pm 0.337}$	-0.333
Rank	–	2.305 ± 0.688	1.567 ± 0.771	0.642 ± 0.053	0.017 ± 0.983	0.619
Uncertainty	–	-2.336 ± 1.344	-1.290 ± 0.341	0.494 ± 0.140	-1.416 ± 0.454	-0.744
Baselines (no imputation)						
Uncertainty	–	1.397 ± 1.699	0.190 ± 0.028	0.541 ± 0.082	0.867 ± 0.198	0.574
Probability	–	0.657 ± 2.145	1.107 ± 0.654	0.626 ± 0.050	0.867 ± 0.198	0.521
Random	–	-0.793 ± 3.159	0.931 ± 0.196	0.555 ± 0.085	-0.086 ± 0.456	0.189

Table 32: Acquisition performance on MIMIC Symile for AUPRC (Lab imputed by ECG), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	4.231 ± 2.318	4.018 ± 1.224	2.093 ± 0.267	6.269 ± 0.383	5.741 ± 1.374	6.317
True KL-Div.	0.924 ± 0.069	0.825 ± 0.017	0.883 ± 0.030	0.717 ± 0.570	0.897 ± 0.019	0.972
True Rank	0.030 ± 0.698	0.643 ± 0.103	0.846 ± 0.041	0.619 ± 1.260	0.351 ± 0.208	0.900
True Uncert.	0.730 ± 0.109	0.793 ± 0.209	0.784 ± 0.034	0.030 ± 0.615	0.740 ± 0.024	0.550
Imputation-based (proposed)						
KL-Div	$\mathbf{0.721 \pm 0.166}$	$\mathbf{1.292 \pm 0.411}$	0.780 ± 0.157	$\mathbf{1.533 \pm 0.215}$	2.267 ± 0.487	**1.779**
Probability	0.532 ± 0.179	1.116 ± 0.598	0.920 ± 0.086	-0.332 ± 0.041	$\mathbf{2.300 \pm 0.464}$	0.751
Rank	0.092 ± 0.485	0.353 ± 0.268	0.372 ± 0.058	-0.032 ± 0.027	-0.641 ± 0.441	0.458
Uncertainty	0.580 ± 0.189	1.080 ± 0.651	$\mathbf{0.930 \pm 0.088}$	-0.549 ± 0.088	0.351 ± 0.131	0.145
Baselines (no imputation)						
Uncertainty	-1.088 ± 1.658	0.451 ± 0.118	0.605 ± 0.085	0.152 ± 0.036	0.645 ± 0.093	0.260
Probability	-0.282 ± 0.309	0.672 ± 0.126	0.594 ± 0.102	0.769 ± 0.144	0.645 ± 0.093	1.112
Random	-0.458 ± 0.636	0.631 ± 0.184	0.659 ± 0.051	0.925 ± 0.146	0.264 ± 0.154	0.556

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	–	8.610 ± 6.961	5.951 ± 0.522	7.534 ± 2.031	12.402 ± 2.525	6.317
True KL-Div.	–	2.009 ± 1.288	0.811 ± 0.344	0.772 ± 0.111	0.910 ± 0.021	0.972
True Rank	–	3.099 ± 2.419	1.137 ± 0.603	0.812 ± 0.056	0.561 ± 0.207	0.900
True Uncert.	–	1.431 ± 1.304	-0.150 ± 0.288	0.555 ± 0.159	0.032 ± 0.114	0.550
Imputation-based (proposed)						
KL-Div	–	$\mathbf{5.709 \pm 4.716}$	$\mathbf{1.662 \pm 0.095}$	$\mathbf{0.667 \pm 0.138}$	1.378 ± 0.274	**1.779**
Probability	–	0.927 ± 0.497	-0.770 ± 0.192	0.647 ± 0.140	$\mathbf{1.420 \pm 0.260}$	0.751
Rank	–	1.988 ± 1.092	0.643 ± 0.448	0.564 ± 0.136	0.788 ± 0.367	0.458
Uncertainty	–	-0.954 ± 0.761	-0.714 ± 0.206	0.591 ± 0.192	-0.013 ± 0.125	0.145
Baselines (no imputation)						
Uncertainty	–	-0.023 ± 0.028	0.078 ± 0.073	0.516 ± 0.137	1.005 ± 0.151	0.260
Probability	–	5.083 ± 4.111	0.941 ± 0.112	0.581 ± 0.069	1.005 ± 0.151	1.112
Random	–	1.804 ± 0.871	0.053 ± 0.419	0.510 ± 0.130	0.616 ± 0.097	0.556

Table 33: Acquisition performance on MIMIC Symile for AUROC (Lab imputed by Image and ECG), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	2.416 ± 0.727	2.670 ± 0.218	3.090 ± 0.355	3.712 ± 0.446	2.174 ± 0.050	4.233
True KL-Div.	1.165 ± 0.267	0.883 ± 0.024	1.050 ± 0.058	1.092 ± 0.080	0.925 ± 0.006	0.981
True Rank	1.373 ± 0.427	0.811 ± 0.024	1.076 ± 0.050	0.904 ± 0.066	0.888 ± 0.008	0.883
True Uncert.	1.008 ± 0.213	0.720 ± 0.056	0.781 ± 0.030	-0.092 ± 0.044	0.675 ± 0.027	0.494
Imputation-based (proposed)						
KL-Div	$\mathbf{1.088 \pm 0.224}$	$\mathbf{0.875 \pm 0.022}$	0.667 ± 0.039	$\mathbf{1.193 \pm 0.091}$	$\mathbf{0.885 \pm 0.008}$	**1.035**
Probability	1.084 ± 0.241	0.647 ± 0.048	0.657 ± 0.039	-0.270 ± 0.085	0.767 ± 0.008	0.390
Rank	-0.190 ± 0.547	0.273 ± 0.051	0.415 ± 0.044	0.448 ± 0.088	0.192 ± 0.017	0.350
Uncertainty	0.981 ± 0.208	0.718 ± 0.054	$\mathbf{0.736 \pm 0.028}$	-0.143 ± 0.044	0.636 ± 0.027	0.448
Baselines (no imputation)						
Uncertainty	0.772 ± 0.126	0.492 ± 0.034	0.627 ± 0.033	0.183 ± 0.030	0.506 ± 0.019	0.469
Probability	0.088 ± 0.147	0.529 ± 0.030	0.376 ± 0.025	0.804 ± 0.069	0.324 ± 0.007	0.521
Random	0.212 ± 0.278	0.404 ± 0.030	0.503 ± 0.030	0.247 ± 0.113	0.439 ± 0.014	0.268

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	2.438 ± 0.113	6.864 ± 1.482	6.398 ± 0.783	4.889 ± 0.635	7.680 ± 0.820	4.233
True KL-Div.	0.887 ± 0.007	1.245 ± 0.526	0.863 ± 0.089	0.809 ± 0.032	0.892 ± 0.057	0.981
True Rank	0.777 ± 0.012	0.383 ± 0.609	0.822 ± 0.095	0.940 ± 0.026	0.855 ± 0.073	0.883
True Uncert.	0.878 ± 0.008	0.148 ± 0.198	0.045 ± 0.084	0.831 ± 0.047	-0.059 ± 0.078	0.494
Imputation-based (proposed)						
KL-Div	$\mathbf{0.887 \pm 0.006}$	$\mathbf{2.529 \pm 0.817}$	$\mathbf{0.914 \pm 0.102}$	0.661 ± 0.036	0.651 ± 0.058	**1.035**
Probability	0.718 ± 0.008	-0.467 ± 0.520	-0.425 ± 0.153	0.445 ± 0.063	$\mathbf{0.743 \pm 0.056}$	0.390
Rank	0.326 ± 0.026	0.909 ± 0.407	0.492 ± 0.087	0.561 ± 0.035	0.079 ± 0.133	0.350
Uncertainty	0.882 ± 0.007	-0.351 ± 0.366	0.040 ± 0.080	$\mathbf{0.838 \pm 0.047}$	0.141 ± 0.068	0.448
Baselines (no imputation)						
Uncertainty	0.693 ± 0.009	-0.022 ± 0.318	0.281 ± 0.070	0.730 ± 0.044	0.425 ± 0.039	0.469
Probability	0.260 ± 0.006	1.476 ± 0.565	0.538 ± 0.131	0.387 ± 0.044	0.425 ± 0.039	0.521
Random	0.334 ± 0.016	-0.383 ± 0.692	0.212 ± 0.120	0.492 ± 0.033	0.216 ± 0.078	0.268

Table 34: Acquisition performance on MIMIC Symile for AUPRC (Lab imputed by Image and ECG), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	1.965 ± 0.497	2.573 ± 0.236	3.714 ± 0.473	2.444 ± 0.174	2.962 ± 0.166	3.748
True KL-Div.	0.961 ± 0.228	0.918 ± 0.033	1.042 ± 0.082	0.985 ± 0.053	0.954 ± 0.013	0.878
True Rank	0.899 ± 0.233	0.896 ± 0.037	1.138 ± 0.112	0.886 ± 0.051	0.922 ± 0.019	0.820
True Uncert.	0.610 ± 0.076	0.741 ± 0.066	0.736 ± 0.075	-0.096 ± 0.042	0.718 ± 0.030	0.482
Imputation-based (proposed)						
KL-Div	$\mathbf{0.914 \pm 0.236}$	$\mathbf{0.921 \pm 0.039}$	0.587 ± 0.056	$\mathbf{1.038 \pm 0.046}$	$\mathbf{0.925 \pm 0.011}$	**0.835**
Probability	0.811 ± 0.100	0.619 ± 0.048	0.677 ± 0.026	-0.376 ± 0.097	0.848 ± 0.014	0.469
Rank	0.516 ± 0.085	0.343 ± 0.047	0.412 ± 0.046	0.355 ± 0.050	0.202 ± 0.019	0.370
Uncertainty	0.662 ± 0.086	0.759 ± 0.070	$\mathbf{0.723 \pm 0.084}$	-0.131 ± 0.030	0.691 ± 0.028	0.482
Baselines (no imputation)						
Uncertainty	0.573 ± 0.099	0.564 ± 0.041	0.601 ± 0.074	0.152 ± 0.016	0.527 ± 0.023	0.485
Probability	0.265 ± 0.173	0.475 ± 0.038	0.424 ± 0.060	0.712 ± 0.038	0.344 ± 0.012	0.518
Random	0.504 ± 0.107	0.449 ± 0.031	0.612 ± 0.063	0.242 ± 0.096	0.471 ± 0.024	0.431

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	3.070 ± 0.202	2.036 ± 0.167	3.622 ± 0.323	5.602 ± 0.520	9.490 ± 1.511	3.748
True KL-Div.	0.890 ± 0.014	0.761 ± 0.095	0.815 ± 0.055	0.824 ± 0.051	0.629 ± 0.127	0.878
True Rank	0.807 ± 0.020	0.452 ± 0.119	0.721 ± 0.057	0.920 ± 0.047	0.560 ± 0.254	0.820
True Uncert.	0.880 ± 0.018	0.171 ± 0.115	0.034 ± 0.040	0.616 ± 0.032	0.411 ± 0.168	0.482
Imputation-based (proposed)						
KL-Div	$\mathbf{0.938 \pm 0.017}$	$\mathbf{0.964 \pm 0.166}$	$\mathbf{0.923 \pm 0.067}$	$\mathbf{0.675 \pm 0.045}$	0.466 ± 0.230	**0.835**
Probability	0.756 ± 0.008	0.182 ± 0.257	-0.400 ± 0.112	0.669 ± 0.034	$\mathbf{0.906 \pm 0.035}$	0.469
Rank	0.368 ± 0.036	0.486 ± 0.142	0.363 ± 0.070	0.418 ± 0.039	0.234 ± 0.154	0.370
Uncertainty	0.932 ± 0.018	0.020 ± 0.100	-0.017 ± 0.041	0.606 ± 0.037	0.573 ± 0.244	0.482
Baselines (no imputation)						
Uncertainty	0.720 ± 0.018	0.158 ± 0.050	0.219 ± 0.024	0.560 ± 0.038	0.781 ± 0.148	0.485
Probability	0.266 ± 0.012	0.860 ± 0.124	0.576 ± 0.066	0.479 ± 0.048	0.781 ± 0.148	0.518
Random	0.375 ± 0.027	0.362 ± 0.172	0.296 ± 0.055	0.464 ± 0.057	0.539 ± 0.133	0.431

Table 35: Acquisition performance on MIMIC Symile for AUROC (ECG imputed by Image), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	3.489 ± 1.656	21.015 ± 5.247	–	–	37.701	28.964
True KL-Div.	0.546 ± 0.073	0.036 ± 0.994	–	–	0.028	0.032
True Rank	0.550 ± 0.379	-2.342 ± 1.928	–	–	-1.276	0.122
True Uncert.	0.648 ± 0.497	-0.099 ± 0.724	–	–	0.379	0.135
Imputation-based (proposed)						
KL-Div	0.084 ± 0.465	-1.318 ± 0.868	–	–	**6.736**	**3.317**
Probability	0.542 ± 0.247	0.096 ± 0.156	–	–	-0.629	0.507
Rank	-0.355 ± 0.092	-1.669 ± 0.386	–	–	1.017	0.062
Uncertainty	$\mathbf{0.672 \pm 0.578}$	0.131 ± 0.351	–	–	-1.407	-0.817
Baselines (no imputation)						
Uncertainty	0.637 ± 0.113	-0.592 ± 0.877	–	–	0.868	0.515
Probability	0.507 ± 0.072	0.236 ± 0.361	–	–	-0.431	-0.192
Random	0.440 ± 0.455	$\mathbf{0.389 \pm 1.478}$	–	–	-1.424	0.484

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	10.077 ± 4.361	2.979 ± 1.139	7.291	–	120.197	28.964
True KL-Div.	0.834 ± 0.029	0.528 ± 0.682	-0.013	–	-1.737	0.032
True Rank	0.645 ± 0.055	1.070 ± 0.717	0.008	–	2.199	0.122
True Uncert.	0.804 ± 0.006	-0.572 ± 0.283	-1.485	–	1.267	0.135
Imputation-based (proposed)						
KL-Div	0.564 ± 0.077	-0.078 ± 0.298	**0.510**	–	**16.719**	**3.317**
Probability	0.450 ± 0.024	-0.870 ± 0.334	-2.693	–	6.656	0.507
Rank	-0.519 ± 0.529	$\mathbf{0.730 \pm 1.632}$	-0.053	–	1.286	0.062
Uncertainty	$\mathbf{0.762 \pm 0.061}$	-0.322 ± 0.369	-1.467	–	-4.089	-0.817
Baselines (no imputation)						
Uncertainty	0.530 ± 0.132	0.704 ± 0.031	0.307	–	1.149	0.515
Probability	0.371 ± 0.024	-1.503 ± 0.841	-1.676	–	1.149	-0.192
Random	0.379 ± 0.012	-1.139 ± 0.014	-0.550	–	5.293	0.484

Table 36: Acquisition performance on MIMIC Symile for AUPRC (ECG imputed by Image), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	1.631	–	100.646 ± 71.555	–	–	33.428
True KL-Div.	0.635	–	2.975 ± 3.762	–	–	1.163
True Rank	0.943	–	-7.003 ± 1.999	–	–	-1.214
True Uncert.	-0.027	–	0.199 ± 0.327	–	–	0.542
Imputation-based (proposed)						
KL-Div	**0.600**	–	$\mathbf{15.994 \pm 7.400}$	–	–	**4.721**
Probability	0.075	–	-10.442 ± 1.040	–	–	-2.263
Rank	-1.052	–	-0.692 ± 4.909	–	–	-0.467
Uncertainty	-0.225	–	-5.787 ± 1.743	–	–	-0.999
Baselines (no imputation)						
Uncertainty	0.456	–	0.652 ± 0.902	–	–	0.692
Probability	0.408	–	0.503 ± 0.861	–	–	0.567
Random	-0.336	–	-7.531 ± 5.766	–	–	-1.297

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	5.642	–	–	–	25.793 ± 2.491	33.428
True KL-Div.	0.855	–	–	–	0.187 ± 0.187	1.163
True Rank	0.784	–	–	–	0.421 ± 0.150	-1.214
True Uncert.	0.919	–	–	–	1.080 ± 0.443	0.542
Imputation-based (proposed)						
KL-Div	0.419	–	–	–	1.872 ± 0.563	**4.721**
Probability	0.491	–	–	–	0.824 ± 0.321	-2.263
Rank	0.340	–	–	–	-0.463 ± 1.077	-0.467
Uncertainty	**0.880**	–	–	–	1.136 ± 0.162	-0.999
Baselines (no imputation)						
Uncertainty	0.685	–	–	–	0.975 ± 0.304	0.692
Probability	0.383	–	–	–	0.975 ± 0.304	0.567
Random	0.360	–	–	–	$\mathbf{2.318 \pm 0.166}$	-1.297

Table 37: Acquisition performance on MIMIC Symile for AUROC (ECG imputed by Lab), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	9.961 ± 1.805	7.528 ± 1.743	7.425	9.429 ± 3.404	42.327	14.888
True KL-Div.	-0.137 ± 1.838	0.978 ± 0.072	0.178	0.356 ± 0.377	0.288	0.405
True Rank	-0.183 ± 0.795	0.976 ± 0.600	0.012	0.962 ± 0.193	-1.285	0.308
True Uncert.	0.271 ± 0.133	0.892 ± 0.394	1.249	0.377 ± 0.354	-0.143	0.399
Imputation-based (proposed)						
KL-Div	-0.756 ± 0.967	0.491 ± 1.880	1.805	$\mathbf{1.782 \pm 0.998}$	$\mathbf{12.001}$	$\mathbf{2.649}$
Probability	0.203 ± 0.341	$\mathbf{2.386 \pm 0.784}$	1.834	1.222 ± 0.693	-15.313	-1.990
Rank	-2.489 ± 1.558	0.182 ± 0.173	-0.177	0.630 ± 0.303	0.104	-0.312
Uncertainty	0.078 ± 0.296	1.509 ± 0.221	$\mathbf{1.945}$	1.538 ± 1.059	1.277	0.335
Baselines (no imputation)						
Uncertainty	0.201 ± 0.098	0.537 ± 0.099	0.628	0.023 ± 0.305	-0.118	0.384
Probability	0.201 ± 0.098	0.474 ± 0.027	0.760	0.650 ± 0.152	0.394	0.074
Random	$\mathbf{1.112 \pm 1.486}$	0.492 ± 0.457	-0.650	0.656 ± 0.102	-1.051	0.387

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	6.604 ± 1.266	2.091 ± 0.714	16.140 ± 11.342	32.694 ± 13.014	14.683	14.888
True KL-Div.	0.800 ± 0.103	1.316 ± 0.077	1.403 ± 0.913	-1.552 ± 1.821	0.425	0.405
True Rank	0.202 ± 0.192	0.974 ± 0.041	2.298 ± 1.714	-0.790 ± 0.663	-0.083	0.308
True Uncert.	0.121 ± 0.119	-0.012 ± 0.256	1.245 ± 0.700	0.285 ± 0.133	-0.292	0.399
Imputation-based (proposed)						
KL-Div	$\mathbf{2.764 \pm 0.736}$	0.746 ± 0.429	0.277 ± 0.403	$\mathbf{4.973 \pm 2.248}$	$\mathbf{2.409}$	$\mathbf{2.649}$
Probability	-2.784 ± 0.997	0.475 ± 0.982	0.657 ± 0.299	-10.662 ± 5.205	2.079	-1.990
Rank	0.337 ± 0.091	0.246 ± 0.433	$\mathbf{1.594 \pm 1.021}$	-3.591 ± 1.582	0.040	-0.312
Uncertainty	-1.728 ± 0.662	0.190 ± 0.470	1.224 ± 0.536	-0.974 ± 2.384	-1.706	0.335
Baselines (no imputation)						
Uncertainty	0.397 ± 0.031	0.312 ± 0.090	1.060 ± 0.408	0.287 ± 0.093	0.512	0.384
Probability	-0.015 ± 0.186	$\mathbf{0.989 \pm 0.490}$	0.043 ± 0.189	-3.264 ± 2.220	0.513	0.074
Random	-0.048 ± 0.167	0.671 ± 0.542	1.329 ± 0.512	2.189 ± 1.764	-0.832	0.387

Table 38: Acquisition performance on MIMIC Symile for AUPRC (ECG imputed by Lab), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	7.894 ± 1.207	7.935 ± 0.606	7.801 ± 1.620	3.681	13.663 ± 0.667	9.827
True KL-Div.	-0.038 ± 1.091	0.843 ± 0.368	-0.530 ± 0.679	0.175	0.815 ± 0.157	0.501
True Rank	0.172 ± 0.009	0.532 ± 0.379	1.309 ± 0.247	0.635	-0.719 ± 0.197	0.354
True Uncert.	0.480 ± 0.166	0.645 ± 0.089	0.762 ± 0.322	0.205	0.039 ± 0.013	0.240
Imputation-based (proposed)						
KL-Div	-1.014 ± 0.482	1.979 ± 1.044	0.596 ± 0.626	0.331	$\mathbf{3.890 \pm 0.009}$	**1.408**
Probability	-0.291 ± 0.867	$\mathbf{2.492 \pm 0.270}$	1.173 ± 0.256	**1.229**	-3.963 ± 0.175	-0.348
Rank	-3.557 ± 1.533	0.231 ± 0.154	-0.653 ± 0.806	0.206	0.003 ± 0.267	-0.331
Uncertainty	0.079 ± 0.427	2.413 ± 0.318	$\mathbf{1.197 \pm 0.340}$	0.817	-0.565 ± 0.214	0.324
Baselines (no imputation)						
Uncertainty	0.398 ± 0.120	0.641 ± 0.487	0.114 ± 0.434	-0.013	0.097 ± 0.015	0.265
Probability	0.398 ± 0.120	0.796 ± 0.109	0.747 ± 0.126	0.861	0.424 ± 0.202	0.581
Random	$\mathbf{0.667 \pm 1.260}$	0.750 ± 0.531	0.520 ± 0.275	0.946	0.091 ± 0.283	0.585

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	5.680 ± 1.134	3.483 ± 1.480	7.752	29.069 ± 4.831	11.309	9.827
True KL-Div.	0.840 ± 0.151	1.316 ± 0.130	0.423	0.558 ± 0.955	0.607	0.501
True Rank	0.341 ± 0.150	1.125 ± 0.278	0.399	-0.106 ± 0.270	-0.148	0.354
True Uncert.	0.037 ± 0.123	-0.790 ± 0.744	0.221	0.157 ± 0.074	0.648	0.240
Imputation-based (proposed)						
KL-Div	$\mathbf{2.195 \pm 0.603}$	0.565 ± 0.242	0.762	$\mathbf{4.430 \pm 1.624}$	0.343	**1.408**
Probability	-2.197 ± 0.763	0.887 ± 0.999	0.932	-4.773 ± 1.949	**1.029**	-0.348
Rank	0.218 ± 0.086	1.061 ± 1.227	0.155	-1.583 ± 0.600	0.613	-0.331
Uncertainty	-1.266 ± 0.476	0.223 ± 0.360	0.606	-0.563 ± 1.118	0.304	0.324
Baselines (no imputation)						
Uncertainty	0.267 ± 0.041	0.055 ± 0.237	-0.024	0.178 ± 0.051	0.936	0.265
Probability	0.038 ± 0.225	$\mathbf{1.274 \pm 0.637}$	**0.945**	-0.603 ± 1.027	0.936	0.581
Random	-0.098 ± 0.129	1.243 ± 0.713	0.239	1.417 ± 0.819	0.076	0.585

Table 39: Acquisition performance on MIMIC Symile for AUROC (ECG imputed by Image and Lab), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	2.634 ± 0.819	4.500 ± 1.001	3.062 ± 0.288	4.806 ± 0.645	2.415 ± 0.090	4.486
True KL-Div.	1.145 ± 0.450	1.172 ± 0.141	0.774 ± 0.037	0.821 ± 0.061	0.897 ± 0.006	0.839
True Rank	1.120 ± 0.318	1.107 ± 0.131	0.858 ± 0.021	0.776 ± 0.057	0.916 ± 0.007	0.836
True Uncert.	1.047 ± 0.381	0.787 ± 0.136	0.632 ± 0.069	0.338 ± 0.071	0.675 ± 0.013	0.537
Imputation-based (proposed)						
KL-Div	$\mathbf{1.120 \pm 0.452}$	$\mathbf{1.200 \pm 0.150}$	$\mathbf{0.771 \pm 0.037}$	$\mathbf{0.812 \pm 0.065}$	$\mathbf{0.894 \pm 0.006}$	**0.838**
Probability	0.933 ± 0.302	0.676 ± 0.106	0.508 ± 0.038	0.111 ± 0.120	0.349 ± 0.014	0.438
Rank	0.126 ± 0.141	0.517 ± 0.086	0.489 ± 0.027	0.507 ± 0.059	0.504 ± 0.009	0.473
Uncertainty	1.063 ± 0.392	0.820 ± 0.140	0.638 ± 0.072	0.358 ± 0.070	0.673 ± 0.013	0.551
Baselines (no imputation)						
Uncertainty	0.785 ± 0.249	0.509 ± 0.088	0.499 ± 0.058	0.468 ± 0.062	0.604 ± 0.009	0.544
Probability	0.097 ± 0.243	0.594 ± 0.105	0.528 ± 0.047	0.435 ± 0.077	0.556 ± 0.012	0.466
Random	0.487 ± 0.212	0.363 ± 0.126	0.463 ± 0.043	0.118 ± 0.177	0.458 ± 0.013	0.427

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	2.675 ± 0.236	3.710 ± 0.511	7.949 ± 2.090	4.600 ± 0.301	8.512 ± 1.512	4.486
True KL-Div.	0.897 ± 0.010	0.549 ± 0.239	0.504 ± 0.303	0.902 ± 0.021	0.724 ± 0.092	0.839
True Rank	0.951 ± 0.011	0.578 ± 0.154	0.661 ± 0.122	0.912 ± 0.025	0.485 ± 0.086	0.836
True Uncert.	0.214 ± 0.031	0.460 ± 0.273	0.414 ± 0.121	0.696 ± 0.035	0.107 ± 0.073	0.537
Imputation-based (proposed)						
KL-Div	$\mathbf{0.865 \pm 0.013}$	0.617 ± 0.239	0.474 ± 0.345	$\mathbf{0.927 \pm 0.022}$	0.704 ± 0.090	**0.838**
Probability	0.188 ± 0.025	0.518 ± 0.303	0.125 ± 0.121	0.178 ± 0.041	$\mathbf{0.789 \pm 0.051}$	0.438
Rank	0.469 ± 0.018	0.458 ± 0.105	$\mathbf{0.886 \pm 0.236}$	0.483 ± 0.030	0.287 ± 0.206	0.473
Uncertainty	0.205 ± 0.029	0.496 ± 0.283	0.436 ± 0.127	0.701 ± 0.034	0.121 ± 0.074	0.551
Baselines (no imputation)						
Uncertainty	0.324 ± 0.015	0.517 ± 0.183	0.569 ± 0.102	0.575 ± 0.036	0.592 ± 0.127	0.544
Probability	0.654 ± 0.014	0.348 ± 0.250	0.256 ± 0.120	0.604 ± 0.017	0.591 ± 0.127	0.466
Random	0.486 ± 0.023	$\mathbf{0.736 \pm 0.238}$	0.351 ± 0.077	0.420 ± 0.038	0.392 ± 0.071	0.427

Table 40: Acquisition performance on MIMIC Symile for AUPRC (ECG imputed by Image and Lab), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	1.578 ± 0.119	3.454 ± 0.465	3.604 ± 0.479	3.295 ± 0.414	2.989 ± 0.181	3.703
True KL-Div.	0.634 ± 0.094	1.070 ± 0.078	0.807 ± 0.038	0.702 ± 0.052	0.904 ± 0.017	0.814
True Rank	0.782 ± 0.060	1.014 ± 0.057	0.873 ± 0.054	0.707 ± 0.076	0.923 ± 0.019	0.800
True Uncert.	0.565 ± 0.081	0.826 ± 0.085	0.531 ± 0.060	0.175 ± 0.056	0.601 ± 0.018	0.431
Imputation-based (proposed)						
KL-Div	$\mathbf{0.621 \pm 0.092}$	$\mathbf{1.087 \pm 0.082}$	$\mathbf{0.815 \pm 0.041}$	$\mathbf{0.702 \pm 0.061}$	$\mathbf{0.903 \pm 0.017}$	**0.812**
Probability	0.577 ± 0.055	0.687 ± 0.059	0.484 ± 0.052	-0.037 ± 0.152	0.382 ± 0.026	0.336
Rank	0.182 ± 0.091	0.481 ± 0.066	0.422 ± 0.074	0.514 ± 0.052	0.469 ± 0.015	0.430
Uncertainty	0.572 ± 0.081	0.851 ± 0.087	0.533 ± 0.061	0.188 ± 0.056	0.602 ± 0.018	0.441
Baselines (no imputation)						
Uncertainty	0.456 ± 0.053	0.596 ± 0.072	0.377 ± 0.054	0.287 ± 0.026	0.521 ± 0.017	0.444
Probability	0.435 ± 0.052	0.502 ± 0.063	0.684 ± 0.050	0.499 ± 0.081	0.660 ± 0.015	0.585
Random	0.374 ± 0.078	0.367 ± 0.062	0.432 ± 0.083	-0.072 ± 0.281	0.430 ± 0.021	0.347

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	2.445 ± 0.148	2.303 ± 0.239	4.150 ± 0.421	5.900 ± 0.609	7.316 ± 0.974	3.703
True KL-Div.	0.856 ± 0.016	0.749 ± 0.133	0.746 ± 0.062	0.952 ± 0.058	0.714 ± 0.062	0.814
True Rank	0.938 ± 0.014	0.690 ± 0.094	0.702 ± 0.072	0.914 ± 0.049	0.461 ± 0.074	0.800
True Uncert.	0.161 ± 0.031	0.098 ± 0.098	0.321 ± 0.066	0.598 ± 0.057	0.438 ± 0.061	0.431
Imputation-based (proposed)						
KL-Div	$\mathbf{0.832 \pm 0.018}$	$\mathbf{0.772 \pm 0.124}$	$\mathbf{0.707 \pm 0.056}$	$\mathbf{0.978 \pm 0.061}$	0.707 ± 0.067	**0.812**
Probability	0.137 ± 0.036	-0.030 ± 0.151	0.176 ± 0.123	0.112 ± 0.108	$\mathbf{0.877 \pm 0.029}$	0.336
Rank	0.426 ± 0.015	0.559 ± 0.123	0.463 ± 0.081	0.378 ± 0.055	0.401 ± 0.059	0.430
Uncertainty	0.154 ± 0.029	0.123 ± 0.086	0.334 ± 0.062	0.603 ± 0.059	0.445 ± 0.061	0.441
Baselines (no imputation)						
Uncertainty	0.279 ± 0.017	0.272 ± 0.065	0.400 ± 0.050	0.436 ± 0.036	0.820 ± 0.050	0.444
Probability	0.674 ± 0.014	0.517 ± 0.139	0.403 ± 0.072	0.655 ± 0.032	0.821 ± 0.050	0.585
Random	0.435 ± 0.016	0.368 ± 0.178	0.362 ± 0.119	0.325 ± 0.048	0.447 ± 0.078	0.347

J RESULTS FOR MIMIC HAIM

Table 41: Acquisition performance on MIMIC HAIM for AUROC, showing G_{full}. Strategies are grouped by category. Best strategy among proposed ones and baselines in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	5.121 ± 1.873	6.802 ± 2.953	4.228 ± 1.374	5.862 ± 0.933	3.435 ± 0.792	4.602
True KL-Div.	0.842 ± 0.154	0.558 ± 0.187	0.673 ± 0.054	0.518 ± 0.093	0.708 ± 0.025	0.608
True Rank	0.853 ± 0.107	0.684 ± 0.056	0.706 ± 0.034	0.676 ± 0.087	0.714 ± 0.041	0.723
True Uncert.	0.429 ± 0.264	0.647 ± 0.107	0.729 ± 0.099	0.332 ± 0.116	0.555 ± 0.057	0.538
Imputation-based (proposed)						
KL-Div	0.827 ± 0.213	0.561 ± 0.142	0.488 ± 0.053	0.505 ± 0.109	0.459 ± 0.074	0.465
Probability	0.318 ± 0.280	$\mathbf{0.657 \pm 0.105}$	0.761 ± 0.133	0.620 ± 0.082	0.578 ± 0.051	0.494
Rank	0.530 ± 0.685	-0.048 ± 0.305	0.311 ± 0.079	0.429 ± 0.150	0.485 ± 0.032	0.391
Uncertainty	0.022 ± 0.463	0.647 ± 0.108	$\mathbf{0.880 \pm 0.241}$	0.503 ± 0.123	0.493 ± 0.056	**0.554**
Baselines (no imputation)						
Uncertainty	0.492 ± 0.209	0.598 ± 0.038	0.486 ± 0.061	0.244 ± 0.163	0.525 ± 0.032	0.452
Probability	0.244 ± 0.302	0.615 ± 0.116	0.550 ± 0.045	$\mathbf{0.641 \pm 0.065}$	$\mathbf{0.580 \pm 0.042}$	0.457
Random	$\mathbf{1.005 \pm 0.367}$	0.143 ± 0.289	0.470 ± 0.045	0.546 ± 0.096	0.510 ± 0.025	0.526

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	4.246 ± 0.416	1.912 ± 0.205	5.429 ± 1.539	2.864 ± 0.716	6.119 ± 2.055	4.602
True KL-Div.	0.609 ± 0.059	0.674 ± 0.148	0.801 ± 0.086	0.775 ± 0.019	-0.083 ± 0.370	0.608
True Rank	0.756 ± 0.027	0.751 ± 0.166	0.898 ± 0.051	0.738 ± 0.022	0.455 ± 0.155	0.723
True Uncert.	0.554 ± 0.047	0.638 ± 0.131	0.330 ± 0.127	0.720 ± 0.019	0.446 ± 0.110	0.538
Imputation-based (proposed)						
KL-Div	0.385 ± 0.058	0.362 ± 0.161	0.418 ± 0.109	0.499 ± 0.046	0.140 ± 0.290	0.465
Probability	0.618 ± 0.041	0.601 ± 0.048	0.636 ± 0.066	$\mathbf{0.591 \pm 0.023}$	-0.435 ± 0.322	0.494
Rank	0.305 ± 0.061	0.422 ± 0.157	0.306 ± 0.194	0.519 ± 0.044	$\mathbf{0.653 \pm 0.050}$	0.391
Uncertainty	0.561 ± 0.042	$\mathbf{0.744 \pm 0.146}$	0.541 ± 0.050	0.587 ± 0.013	0.558 ± 0.102	**0.554**
Baselines (no imputation)						
Uncertainty	0.460 ± 0.038	0.511 ± 0.063	0.396 ± 0.095	0.539 ± 0.023	0.272 ± 0.185	0.452
Probability	$\mathbf{0.619 \pm 0.043}$	0.508 ± 0.086	$\mathbf{0.754 \pm 0.050}$	0.575 ± 0.025	-0.513 ± 0.421	0.457
Random	0.482 ± 0.022	0.703 ± 0.129	0.534 ± 0.038	0.553 ± 0.009	0.319 ± 0.136	0.526

Table 42: Acquisition performance on MIMIC HAIM for AUPRC, showing G_{full}. Strategies are grouped by category. Best strategy among proposed ones and baselines in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	1.490 ± 0.249	3.181 ± 0.283	3.084 ± 0.331	2.738 ± 0.363	2.782 ± 0.349	3.087
True KL-Div.	0.967 ± 0.098	0.652 ± 0.076	0.435 ± 0.167	0.847 ± 0.065	0.717 ± 0.030	0.662
True Rank	1.014 ± 0.085	0.721 ± 0.068	0.541 ± 0.112	0.958 ± 0.097	0.769 ± 0.031	0.754
True Uncert.	0.842 ± 0.068	0.592 ± 0.081	0.605 ± 0.047	0.425 ± 0.084	0.611 ± 0.057	0.587
Imputation-based (proposed)						
KL-Div	0.889 ± 0.097	0.578 ± 0.054	0.301 ± 0.163	0.655 ± 0.110	0.518 ± 0.030	0.516
Probability	0.584 ± 0.111	0.672 ± 0.041	$\mathbf{0.807 \pm 0.075}$	0.480 ± 0.018	0.754 ± 0.017	0.686
Rank	0.890 ± 0.120	0.337 ± 0.081	0.247 ± 0.115	$\mathbf{0.706 \pm 0.052}$	$\mathbf{0.474 \pm 0.023}$	0.456
Uncertainty	0.799 ± 0.056	0.540 ± 0.082	0.680 ± 0.071	0.353 ± 0.047	0.584 ± 0.030	0.570
Baselines (no imputation)						
Uncertainty	0.920 ± 0.131	0.556 ± 0.082	0.409 ± 0.085	0.365 ± 0.011	0.487 ± 0.053	0.479
Probability	0.480 ± 0.112	$\mathbf{0.713 \pm 0.046}$	0.755 ± 0.054	0.589 ± 0.053	0.752 ± 0.017	**0.701**
Random	$\mathbf{0.941 \pm 0.107}$	0.404 ± 0.060	0.455 ± 0.057	0.452 ± 0.180	0.546 ± 0.024	0.561

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	4.751 ± 0.276	1.878 ± 0.237	2.730 ± 0.219	3.910 ± 0.892	4.329 ± 0.483	3.087
True KL-Div.	0.374 ± 0.137	0.688 ± 0.130	0.660 ± 0.196	0.699 ± 0.035	0.575 ± 0.039	0.662
True Rank	0.636 ± 0.071	0.725 ± 0.153	0.861 ± 0.144	0.683 ± 0.035	0.630 ± 0.042	0.754
True Uncert.	0.599 ± 0.059	0.430 ± 0.127	0.505 ± 0.173	0.732 ± 0.031	0.530 ± 0.043	0.587
Imputation-based (proposed)						
KL-Div	0.248 ± 0.124	0.573 ± 0.133	0.413 ± 0.218	0.399 ± 0.089	$\mathbf{0.588 \pm 0.043}$	0.516
Probability	$\mathbf{0.808 \pm 0.057}$	0.764 ± 0.115	0.737 ± 0.030	$\mathbf{0.736 \pm 0.025}$	0.518 ± 0.040	0.686
Rank	0.167 ± 0.117	0.216 ± 0.119	0.494 ± 0.147	0.490 ± 0.079	0.538 ± 0.024	0.456
Uncertainty	0.641 ± 0.052	0.450 ± 0.120	0.561 ± 0.092	0.621 ± 0.030	0.468 ± 0.030	0.570
Baselines (no imputation)						
Uncertainty	0.372 ± 0.063	0.274 ± 0.111	0.359 ± 0.220	0.496 ± 0.042	0.550 ± 0.036	0.479
Probability	0.795 ± 0.076	$\mathbf{0.806 \pm 0.109}$	$\mathbf{0.855 \pm 0.043}$	0.725 ± 0.021	0.536 ± 0.033	**0.701**
Random	0.504 ± 0.042	0.606 ± 0.151	0.572 ± 0.086	0.569 ± 0.022	0.556 ± 0.022	0.561

Table 43: Acquisition performance on MIMIC HAIM for AUROC (Image imputed by Lab), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	4.447 ± 3.050	3.381 ± 1.055	2.055 ± 0.185	6.068 ± 1.280	1.710 ± 0.036	3.810
True KL-Div.	1.002 ± 0.142	0.425 ± 0.184	0.623 ± 0.061	0.299 ± 0.113	0.687 ± 0.022	0.476
True Rank	0.902 ± 0.059	0.704 ± 0.044	0.667 ± 0.035	0.575 ± 0.139	0.698 ± 0.010	0.671
True Uncert.	0.908 ± 0.185	0.768 ± 0.035	0.641 ± 0.022	0.301 ± 0.207	0.709 ± 0.008	0.610
Imputation-based (proposed)						
KL-Div	0.686 ± 0.060	0.579 ± 0.031	0.463 ± 0.054	0.412 ± 0.085	0.539 ± 0.022	0.416
Probability	0.514 ± 0.051	$\mathbf{0.773 \pm 0.153}$	$\mathbf{0.560 \pm 0.026}$	$\mathbf{0.768 \pm 0.113}$	0.590 ± 0.010	0.445
Rank	$\mathbf{1.530 \pm 0.782}$	0.045 ± 0.280	0.418 ± 0.037	0.203 ± 0.164	0.450 ± 0.021	0.497
Uncertainty	0.710 ± 0.087	0.730 ± 0.091	0.548 ± 0.029	0.597 ± 0.215	0.587 ± 0.012	**0.625**
Baselines (no imputation)						
Uncertainty	0.512 ± 0.247	0.635 ± 0.040	0.491 ± 0.032	0.086 ± 0.287	0.587 ± 0.015	0.438
Probability	0.109 ± 0.371	0.759 ± 0.146	0.533 ± 0.026	0.739 ± 0.095	$\mathbf{0.592 \pm 0.010}$	0.392
Random	1.112 ± 0.371	0.483 ± 0.041	0.509 ± 0.015	0.556 ± 0.138	0.542 ± 0.008	0.582

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	3.257 ± 0.286	2.139 ± 0.317	3.951 ± 0.558	1.968 ± 0.068	9.124 ± 3.961	3.810
True KL-Div.	0.440 ± 0.063	0.714 ± 0.312	0.734 ± 0.096	0.775 ± 0.021	-0.936 ± 0.646	0.476
True Rank	0.671 ± 0.017	0.808 ± 0.318	0.898 ± 0.066	0.711 ± 0.015	0.081 ± 0.265	0.671
True Uncert.	0.558 ± 0.042	0.625 ± 0.057	0.456 ± 0.097	0.734 ± 0.020	0.400 ± 0.221	0.610
Imputation-based (proposed)						
KL-Div	0.321 ± 0.083	0.458 ± 0.273	0.342 ± 0.129	0.552 ± 0.027	-0.187 ± 0.576	0.416
Probability	$\mathbf{0.698 \pm 0.059}$	0.544 ± 0.099	0.693 ± 0.063	$\mathbf{0.600 \pm 0.020}$	-1.295 ± 0.522	0.445
Rank	0.233 ± 0.093	0.339 ± 0.349	0.468 ± 0.090	0.498 ± 0.017	$\mathbf{0.783 \pm 0.075}$	0.497
Uncertainty	0.620 ± 0.047	0.668 ± 0.165	0.581 ± 0.057	0.591 ± 0.015	0.613 ± 0.206	**0.625**
Baselines (no imputation)						
Uncertainty	0.434 ± 0.052	0.568 ± 0.074	0.394 ± 0.109	0.557 ± 0.023	0.114 ± 0.373	0.438
Probability	0.693 ± 0.057	0.541 ± 0.162	$\mathbf{0.790 \pm 0.054}$	0.589 ± 0.018	-1.427 ± 0.750	0.392
Random	0.482 ± 0.030	$\mathbf{0.941 \pm 0.176}$	0.518 ± 0.049	0.558 ± 0.010	0.119 ± 0.262	0.582

Table 44: Acquisition performance on MIMIC HAIM for AUPRC (Image imputed by Lab), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	1.490 ± 0.249	3.177 ± 0.420	3.038 ± 0.399	2.586	1.812 ± 0.037	2.790
True KL-Div.	0.967 ± 0.098	0.523 ± 0.091	0.372 ± 0.195	0.717	0.664 ± 0.033	0.579
True Rank	1.014 ± 0.085	0.632 ± 0.084	0.498 ± 0.128	0.773	0.685 ± 0.019	0.706
True Uncert.	0.842 ± 0.068	0.790 ± 0.040	0.624 ± 0.054	0.575	0.776 ± 0.014	0.660
Imputation-based (proposed)						
KL-Div	0.889 ± 0.097	0.540 ± 0.066	0.260 ± 0.194	0.456	0.586 ± 0.031	0.478
Probability	0.584 ± 0.111	0.674 ± 0.058	$\mathbf{0.796 \pm 0.087}$	0.455	0.710 ± 0.011	0.662
Rank	0.890 ± 0.120	0.288 ± 0.118	0.229 ± 0.139	**0.650**	0.459 ± 0.026	0.416
Uncertainty	0.799 ± 0.056	0.683 ± 0.085	0.700 ± 0.081	0.263	0.665 ± 0.022	0.615
Baselines (no imputation)						
Uncertainty	0.920 ± 0.131	0.700 ± 0.096	0.448 ± 0.096	0.387	0.651 ± 0.025	0.553
Probability	0.480 ± 0.112	$\mathbf{0.714 \pm 0.071}$	0.749 ± 0.063	0.483	$\mathbf{0.715 \pm 0.011}$	**0.686**
Random	$\mathbf{0.941 \pm 0.107}$	0.429 ± 0.088	0.464 ± 0.067	0.590	0.595 ± 0.012	0.615

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	4.700 ± 0.289	1.725 ± 0.306	2.830 ± 0.264	2.510 ± 0.170	4.037 ± 0.525	2.790
True KL-Div.	0.054 ± 0.147	0.651 ± 0.392	0.621 ± 0.266	0.731 ± 0.039	0.493 ± 0.045	0.579
True Rank	0.463 ± 0.068	0.834 ± 0.425	0.903 ± 0.193	0.661 ± 0.030	0.601 ± 0.047	0.706
True Uncert.	0.490 ± 0.057	0.666 ± 0.066	0.495 ± 0.235	0.726 ± 0.038	0.615 ± 0.050	0.660
Imputation-based (proposed)						
KL-Div	0.064 ± 0.176	0.384 ± 0.277	0.385 ± 0.295	0.540 ± 0.044	$\mathbf{0.680 \pm 0.038}$	0.478
Probability	0.920 ± 0.072	0.556 ± 0.160	0.718 ± 0.033	0.725 ± 0.030	0.485 ± 0.043	0.662
Rank	-0.025 ± 0.160	0.216 ± 0.282	0.456 ± 0.198	0.478 ± 0.034	0.518 ± 0.025	0.416
Uncertainty	0.655 ± 0.077	0.656 ± 0.200	0.574 ± 0.124	0.638 ± 0.037	0.519 ± 0.034	0.615
Baselines (no imputation)						
Uncertainty	0.293 ± 0.089	0.529 ± 0.084	0.416 ± 0.296	0.544 ± 0.042	0.643 ± 0.024	0.553
Probability	$\mathbf{0.944 \pm 0.092}$	0.671 ± 0.071	$\mathbf{0.850 \pm 0.058}$	$\mathbf{0.732 \pm 0.022}$	0.523 ± 0.031	**0.686**
Random	0.462 ± 0.059	$\mathbf{0.958 \pm 0.269}$	0.576 ± 0.117	0.580 ± 0.019	0.553 ± 0.031	0.615

Table 45: Acquisition performance on MIMIC HAIM for AUROC (Lab imputed by Image), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	6.244 ± 0.851	12.503 ± 7.484	11.469 ± 3.810	5.604 ± 1.451	5.899 ± 1.530	6.851
True KL-Div.	0.575 ± 0.311	0.780 ± 0.402	0.842 ± 0.049	0.792 ± 0.088	0.737 ± 0.054	0.770
True Rank	0.773 ± 0.299	0.649 ± 0.136	0.836 ± 0.030	0.802 ± 0.076	0.736 ± 0.102	0.796
True Uncert.	-0.370 ± 0.158	0.445 ± 0.274	1.020 ± 0.440	0.371 ± 0.065	0.334 ± 0.084	0.416
Imputation-based (proposed)						
KL-Div	$\mathbf{1.063 \pm 0.606}$	0.531 ± 0.397	0.572 ± 0.156	0.622 ± 0.223	0.345 ± 0.175	0.519
Probability	-0.010 ± 0.798	0.463 ± 0.079	1.433 ± 0.393	0.435 ± 0.086	0.560 ± 0.129	**0.551**
Rank	-1.136 ± 0.350	-0.202 ± 0.707	-0.044 ± 0.241	$\mathbf{0.712 \pm 0.244}$	0.535 ± 0.072	0.180
Uncertainty	-1.124 ± 0.961	0.508 ± 0.248	$\mathbf{1.986 \pm 0.840}$	0.385 ± 0.063	0.359 ± 0.123	0.492
Baselines (no imputation)						
Uncertainty	0.460 ± 0.449	$\mathbf{0.535 \pm 0.076}$	0.467 ± 0.283	0.441 ± 0.053	0.438 ± 0.063	0.458
Probability	0.470 ± 0.594	0.375 ± 0.158	0.608 ± 0.200	0.519 ± 0.069	$\mathbf{0.563 \pm 0.105}$	0.511
Random	0.826 ± 0.879	-0.425 ± 0.748	0.337 ± 0.193	0.534 ± 0.143	0.464 ± 0.056	0.435

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	5.483 ± 0.647	1.730 ± 0.268	9.123 ± 5.217	7.345 ± 3.084	3.113 ± 0.375	6.851
True KL-Div.	0.821 ± 0.038	0.641 ± 0.142	0.969 ± 0.173	0.774 ± 0.071	0.771 ± 0.016	0.770
True Rank	0.861 ± 0.023	0.706 ± 0.194	0.896 ± 0.088	0.873 ± 0.035	0.830 ± 0.017	0.796
True Uncert.	0.550 ± 0.096	0.648 ± 0.244	0.017 ± 0.360	0.652 ± 0.013	0.492 ± 0.038	0.416
Imputation-based (proposed)						
KL-Div	0.465 ± 0.074	0.286 ± 0.213	0.608 ± 0.194	0.233 ± 0.147	0.467 ± 0.028	0.519
Probability	0.518 ± 0.032	0.646 ± 0.037	0.492 ± 0.164	0.543 ± 0.113	0.425 ± 0.017	**0.551**
Rank	0.395 ± 0.065	0.489 ± 0.114	-0.098 ± 0.659	$\mathbf{0.626 \pm 0.318}$	$\mathbf{0.522 \pm 0.033}$	0.180
Uncertainty	0.487 ± 0.070	$\mathbf{0.805 \pm 0.241}$	0.441 ± 0.090	0.570 ± 0.037	0.503 ± 0.028	0.492
Baselines (no imputation)						
Uncertainty	0.492 ± 0.057	0.466 ± 0.100	0.401 ± 0.220	0.454 ± 0.058	0.430 ± 0.020	0.458
Probability	$\mathbf{0.526 \pm 0.049}$	0.482 ± 0.103	$\mathbf{0.664 \pm 0.114}$	0.508 ± 0.151	0.400 ± 0.018	0.511
Random	0.482 ± 0.033	0.512 ± 0.145	0.575 ± 0.056	0.530 ± 0.018	0.519 ± 0.024	0.435

Table 46: Acquisition performance on MIMIC HAIM for AUPRC (Lab imputed by Image), showing G_{full}. Strategies are grouped by category. Best strategy among proposed and baseline methods in bold for each column.

Strategy	Fracture	Enl. Card.	Consolidation	Atelectasis	Edema	Mean
Upper Bounds (for reference)						
Oracle	–	3.188 ± 0.307	3.312 ± 0.007	2.814 ± 0.615	4.169 ± 0.490	4.011
True KL-Div.	–	0.884 ± 0.047	0.749 ± 0.053	0.912 ± 0.012	0.792 ± 0.043	0.776
True Rank	–	0.881 ± 0.081	0.758 ± 0.195	1.051 ± 0.050	0.889 ± 0.035	0.811
True Uncert.	–	0.237 ± 0.075	0.508 ± 0.005	0.350 ± 0.067	0.376 ± 0.071	0.467
Imputation-based (proposed)						
KL-Div	–	0.647 ± 0.098	0.508 ± 0.078	$\mathbf{0.755 \pm 0.079}$	0.419 ± 0.032	0.489
Probability	–	0.668 ± 0.054	$\mathbf{0.858 \pm 0.163}$	0.492 ± 0.022	$\mathbf{0.816 \pm 0.022}$	0.723
Rank	–	0.424 ± 0.082	0.335 ± 0.044	0.734 ± 0.076	0.495 ± 0.042	0.485
Uncertainty	–	0.282 ± 0.094	0.580 ± 0.160	0.398 ± 0.022	0.469 ± 0.034	0.461
Baselines (no imputation)						
Uncertainty	–	0.298 ± 0.039	0.214 ± 0.113	0.354 ± 0.001	0.254 ± 0.039	0.294
Probability	–	$\mathbf{0.712 \pm 0.031}$	0.782 ± 0.104	0.642 ± 0.002	0.806 ± 0.028	**0.726**
Random	–	0.361 ± 0.069	0.406 ± 0.089	0.383 ± 0.287	0.478 ± 0.045	0.472

Strategy	Cardiomegaly	Lung Lesion	Lung Opacity	Pneumonia	Pneumothorax	Mean
Upper Bounds (for reference)						
Oracle	4.828 ± 0.574	1.969 ± 0.352	2.430 ± 0.426	8.578 ± 2.467	4.816 ± 0.981	4.011
True KL-Div.	0.854 ± 0.056	0.711 ± 0.032	0.776 ± 0.034	0.595 ± 0.055	0.712 ± 0.014	0.776
True Rank	0.895 ± 0.034	0.660 ± 0.093	0.735 ± 0.039	0.756 ± 0.125	0.678 ± 0.081	0.811
True Uncert.	0.762 ± 0.089	0.289 ± 0.175	0.537 ± 0.099	0.752 ± 0.052	0.389 ± 0.036	0.467
Imputation-based (proposed)						
KL-Div	0.524 ± 0.084	0.687 ± 0.135	0.495 ± 0.131	-0.069 ± 0.185	0.434 ± 0.051	0.489
Probability	$\mathbf{0.640 \pm 0.029}$	$\mathbf{0.889 \pm 0.138}$	0.794 ± 0.073	$\mathbf{0.773 \pm 0.051}$	$\mathbf{0.574 \pm 0.078}$	0.723
Rank	0.455 ± 0.084	0.216 ± 0.127	0.607 ± 0.038	0.533 ± 0.382	0.570 ± 0.049	0.485
Uncertainty	0.621 ± 0.069	0.326 ± 0.134	0.523 ± 0.042	0.566 ± 0.018	0.381 ± 0.036	0.461
Baselines (no imputation)						
Uncertainty	0.489 ± 0.062	0.121 ± 0.130	0.187 ± 0.017	0.333 ± 0.042	0.396 ± 0.034	0.294
Probability	0.571 ± 0.058	0.887 ± 0.166	$\mathbf{0.870 \pm 0.029}$	0.703 ± 0.060	0.559 ± 0.076	**0.726**
Random	0.568 ± 0.050	0.395 ± 0.110	0.561 ± 0.044	0.535 ± 0.077	0.560 ± 0.029	0.472