TEST-TIME VIEW SELECTION FOR MULTI-MODAL DECISION MAKING

Eeshaan Jain[1,*], **Johann Wenckstern**[1], **Benedikt von Querfurth**[1], **Charlotte Bunne**[1,2]

[1]School of Computer and Communication Science, EPFL
[2]School of Life Sciences, EPFL

ABSTRACT

The clinical routine has access to an ever-expanding repertoire of diagnostic tests, ranging from routine imaging to sophisticated molecular profiling technologies. Foundation models have recently emerged as powerful tools for extracting and integrating diagnostic information from these diverse clinical tests, advancing the idea of comprehensive patient digital twins. However, it remains unclear how to select and design tests that ensure foundation models can extract the necessary information for accurate diagnosis. We introduce MAVIS (Multi-modal Active VIew Selection), a reinforcement learning framework that unifies modality selection and feature selection into a single decision process. By leveraging foundation models, MAVIS dynamically determines which diagnostic tests to perform and in what sequence, adapting to individual patient characteristics. Experiments on real-world datasets across multiple clinical tasks demonstrate that MAVIS outperforms conventional approaches in both diagnostic accuracy and uncertainty reduction, while reducing testing costs by over 80%, suggesting a promising direction for optimizing clinical workflows through intelligent test design and selection.

1 INTRODUCTION

Recent advances in medical diagnostics have provided clinicians with an unprecedented wealth of data modalities to inform patient care decisions. Such tests, however, might strongly differ in costs, effort, and availability. In cancer diagnostics, for example, genomic and pathological profiling of a patient's tumor biopsy allows to stage tumors, provide prognostic information, and guide therapeutic decision making. While routine and cost-effective hematoxylin and eosin (H&E) staining of a patient's tumor biopsy has clear clinical utility, the complexity of tumor tissues may require a more comprehensive assessment of molecular details to fulfill the promise of precision oncology, and thus more elaborate profiling assays such as spatial proteomics (SP) are required (Fig. 1a) (Gong et al., 2024; Nordmann et al., 2024; de Souza et al., 2024).

The emergence of foundation models (FM) capable of processing and integrating these diverse data types has created new opportunities for more accurate diagnostics, treatment selection, and prognostic predictions (Bunne et al., 2024). Trained on large collections of datasets often spanning different modalities, FMs allow analyzing patient-derived data in the context of previous measurements, and as such have the potential to complement missing or incomplete information of a patient's molecular profile. The power of these models introduces a key practical consideration in clinical settings:

> *Given such powerful foundation models, how can we optimally select and sequence diagnostic tests while balancing clinical utility against cost and time constraints?*

The challenge is particularly acute in molecular profiling technologies such as spatial proteomics (Black et al., 2021; Giesen et al., 2014; Lewis et al., 2021; Danenberg et al., 2022), where each additional protein marker measured significantly increases both cost and experimental complexity. While comprehensive molecular profiling would provide the most complete picture of a patient's condition, practical constraints necessitate strategic selection of the most informative markers and rational experimental design (Chaloner & Verdinelli, 1995). This selection process must account not only for the immediate diagnostic value but also for the downstream impact on treatment decisions and prognostic assessments (Fig. 1b).

* Correspondence to eeshaan.jain@epfl.ch.

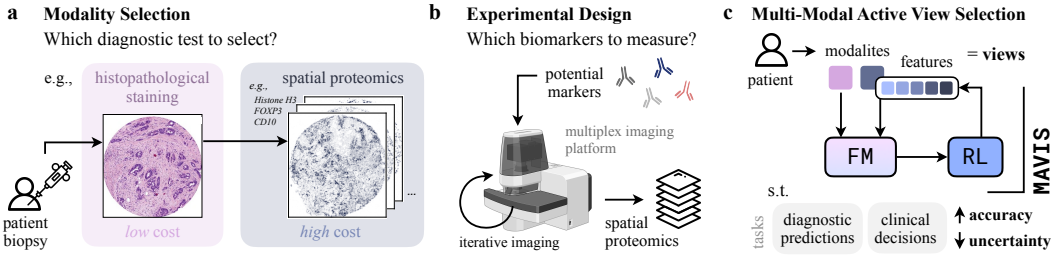

Figure 1: MAVIS tackles problems such as **a.** modality selection, i.e., deciding on an optimal diagnostic test to choose between low-cost H&E staining and high-cost spatial proteomics, **b.** experimental design for biomarker selection, through **c.** integrating foundation models and multi-modal RL for test-time decision making and active view selection.

Traditional methodologies frame the problem of the selection of diagonostic platforms (referred to here as *modalities*) and selection of specific measurement features within each diagnostic platform (referred to here as *features*), such as markers used within a spatial proteomic experiment, as separate optimization problems. However, one might argue that both, different modalities as well as different features of a modality might simply provide different *views* on a sample. Besides selecting different *modalities* or *views*, it is crucial to determine their optimal ordering. For example, restricting the series of diagnostic tests conducted for a patient to cost-efficient routine analyses and only moving to more sophisticated, labor-intensive, and expensive test if required, significantly reduces healthcare costs and increases efficiency. However, determining an optimal sequence of diagnostic tests is challenging due to the (i) the heterogeneity present in both experimental conditions and patient cohorts, and (ii) the absence of ground truth test sequences in historically collected datasets.

In this paper, we introduce **M**ulti-modal **A**ctive **Vi**ew **S**election (MAVIS), a novel reinforcement learning (RL)-based framework for active view selection at inference-time in clinical diagnostics. MAVIS iteratively identifies the most informative diagnostic tests and features, aiming to maximize accuracy while minimizing uncertainty across diverse downstream clinical tasks. MAVIS builds on foundation models as a basis for informed decision-making, combining their predictive power with reinforcement learning to dynamically select the most informative diagnostic tests for each patient (Fig. 1c). Our key contributions include:

Unified framework for modality and feature selection. Contrary to prior work, we propose a single framework for selecting diagnostic platforms (*modalities*) and their corresponding measurement features by phrasing the problem as *view selection task* across multiple modalities. Within this, we jointly tackle two critical questions: (i) For which patient is H&E staining alone inadequate for diagnostic decision-making, thereby necessitating spatial proteomics measurements? (ii) What is the optimal sequence of markers to measure iteratively to strengthen diagnostic precision? This allows us to explicitly account for cost-benefit trade-offs in diagnostic test selection.

Novel RL-based method for selection at test-time. We propose MAVIS, a RL framework, which learns to iteratively select the sequence of views to be chosen to enhance diagnostic prediction, leveraging historical experimental data. Unlike prior approaches, MAVIS is trained in an environment simulating test-time conditions, and hence it does not have prior access to results of unseen views, nor requires supervision through ground-truth preference orders or view sequences.

Guiding test-time selection through foundation models. Recent advances in FMs for pathology (Chen et al., 2024) and spatial proteomics (Wenckstern et al., 2025) have demonstrated excellent zero-shot performance capabilities in generating high-quality tissue representations and in downstream tasks. We leverage these pre-trained FMs within MAVIS for guiding modality and view selection. As such, MAVIS is the first framework that utilizes multiple FMs as backbones for decision making with without the need for fine-tuning or additional supervision at test time.

2 Background

Foundation models in clinical diagnostics. FMs have gained significant attention in the medical imaging and computational biology communities for their ability to learn powerful representations from diverse datasets (Chen et al., 2024; Wenckstern et al., 2025; Vorontsov et al., 2024; Xu et al.,

2024). These models have shown remarkable downstream performance without additional supervision. In this work, we use UNI2 (Chen et al., 2024) as the underlying FM for H&E stained images, a model with strong zero-shot and few-shot performance on effectively classifying and segmenting histopathology images without requiring extensive task-specific training data. For spatial proteomics, we utilize VirTues (Wenckstern et al., 2025) to encode multiplexed images, since it is adept at encoding multiple marker measurements into a single embedding, capturing complex relationships within an evolving representation of a potentially growing number of selected markers.

While foundation models demonstrate impressive capabilities, their real-world deployment often faces challenges due to distribution shifts. Hübotter et al. (2024) address this through combining test-time training (Sun et al., 2020) with active learning (Settles, 2009), which allows models to adapt to test samples during inference through self-supervised learning objectives, without requiring additional labeled data. Their approach actively selects samples to reduce uncertainty but requires prior knowledge of the complete data space—a limitation in clinical settings. Gupte et al. (2024) propose an uncertainty-based active learning strategies for improving the performance of foundation models on downstream tasks, though focusing on sample selection rather than feature acquisition.

Reinforcement learning. RL has proven effective for a variety of sequential decision-making problems (Sutton, 2018), from game-playing agents (Vinyals et al., 2019) to autonomous vehicles (Kiran et al., 2021). More recently, reinforcement learning from human feedback (RLHF) has been employed to refine reasoning in large language models (LLMs), guiding them toward generating more coherent and contextually appropriate responses (Kumar et al., 2024; Guo et al., 2025; Shao et al., 2024). In the context of clinical diagnostics (Ling et al., 2017; Coronato et al., 2020; Peng et al., 2018), RL offers a principled way to iteratively decide which diagnostic analysis or features to acquire at test-time in the clinic, where fine-tuning and supervision are not possible.

Formally, the RL problem can be framed as a Markov decision process (MDP) $(\mathcal{S}, \mathcal{A}, P, c, \gamma)$, with $\mathcal{S}$ as the set of possible states (partial test results), $\mathcal{A}$ being the set of actions (which test/marker to select), $P : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ being the transition function defining how states evolve after taking actions, $c : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ as the reward function, and γ as the discount factor. The goal of RL is to learn a policy $\pi_\theta : \mathcal{S} \to \Delta(\mathcal{A})$ mapping states to a probability distribution over the discrete set $\mathcal{A}$. With τ being the sequence of $((s_t, a_t, c_t))_{t=0}^{T}$, the goal of policy optimization is to maximize the expected γ-discounted cumulative return $J(\theta) = \mathbb{E}_{\tau \sim P_\pi(\tau)} \left[\sum_{t=0}^{T} \gamma^t c(s_t, a_t)\right]$, and $\pi^* = \operatorname{argmax}_\pi J(\theta)$. Proximal policy optimization (PPO) (Schulman et al., 2017) is an effective model-free policy gradient algorithm, which learns a policy π_θ and a value function $V_\theta : \mathcal{S} \to \mathbb{R}$. After collecting a set of new episodes $\mathcal{E}$, using the policy prior to the update step, PPO updates the policy and value networks by optimizing
$L(\theta) = \mathbb{E}_{(s_t,a_t)\sim\mathcal{E}} \left[\min\left(r_t(\theta)\hat{A}_t, \text{clip}\left(r_t(\theta), 1-\epsilon, 1+\epsilon\right)\hat{A}_t\right) - \lambda_1 \left(V_\theta(s_t) - R_t\right)^2 + \lambda_2 H(\pi_\theta)\right]$,
where $H(\pi_\theta) = -\sum_{a\in\mathcal{A}} \pi_\theta(a|s_t) \log \pi_\theta(a|s_t)$, and $\lambda_1, \lambda_2, \epsilon$ are hyperparameters, and the advantage $\hat{A}_t$ is calculated using an estimation scheme known as generalized advantage estimator (GAE) (Schulman et al., 2015).
For further background, see Appendix **??**.

3 METHOD

Notation. In our setting, we assume all patients have access to the same views, *i.e.*, modalities and corresponding features. We denote the set of modalities by $\mathcal{M}$, and set of views belonging to modality $m_k \in \mathcal{M}$ as $\mathcal{V}_k$. For each patient $p \in \mathcal{P}$, view v belonging to modality m has an associated measurement denoted as $\mathbf{X}_m^p[v]$, and we overload the notation to denote $\mathbf{X}_m^p[\mathcal{T}] = \{\mathbf{X}_m^p[v] \,|\, v \in \mathcal{T}\}$. We denote the foundation model for modality m as $F_m : \{\mathbf{X}_m[\mathcal{T}] \mid \mathcal{T} \in \mathcal{P}(\mathcal{V}_m) \setminus \emptyset\} \to \mathbb{R}^{d_m}$, which gives an embedding $Z_m[\mathcal{T}] \in \mathbb{R}^{d_m}$ for any non-empty subset $\mathcal{T} \subseteq \mathcal{V}_m$. For brevity, we simplify the notation to denote $F_m(\mathcal{T}) \equiv F_m(\mathbf{X}_m[\mathcal{T}])$. For a given task with labels $\mathcal{Y}$, the downstream randomized classifier for modality m is given as $h_m : \mathbb{R}^{d_m} \to \Delta\mathcal{Y}$, which is trained on historically collected data $D_{\text{train}} = \{(Z^i_{m,\mathcal{V}_m}, y^i)\}_{i=1}^{N_{\text{train}}}$, which lacks sequential or preferential information among views, and is highly heterogeneous across both patient and task cohorts.

3.1 MAVIS FOR UNI-MODAL VIEW SELECTION

We define the challenge of uni-modal view selection as follows: What is the optimal order of views to select given a patient? For example, in spatial proteomics, marker measurements are typically conducted alongside a base marker, which serves purposes such as cellular alignment or segmentation. Since each further marker measurement incurs an incremental cost, determining the most efficient ordering of acquisitions and *minimal diagnostic screening* for a patient are critical to balance diagnostic accuracy and resource efficiency. Unlike modalities, which represent distinct experimental setups and are typically assessed independently, different views (in this case, markers) present a unique challenge to deal with, as decisions are often informed by their collective rather than individual contributions. This distinction makes our setting slightly different from standard active feature acquisition, where features are usually assumed to be independently informative and also from active test-time fine-tuning, which involves selecting adaptation points to refine model predictions but requires prior access to the data space and test-time training. In contrast, we design MAVIS to operate in a purely inference-time setting, sequentially acquiring markers without additional model updates. We now formalize uni-modal view selection tasks as follows: Since we consider a single modality, we drop the subscript m in this description. Let $\mathcal{V} = (V_1, \ldots, V_n)$ be the set of views for a modality, and let the foundation model be F, with the downstream classifier h.

Definition 1 (Selection rule)**.** A selection rule is a function $\sigma^\ell : \mathcal{P} \to S_{|\mathcal{V}|}$, where $S_{|\mathcal{V}|}$ is the permutation group of $|\mathcal{V}|$ elements, interpreted as the set of orders on $\mathcal{V}$, such that for each patient $p \in \mathcal{P}$, $\sigma^\ell(p)$ is an order of views in $\mathcal{V}$ with $\sigma^\ell(p)(1) = V_\ell$.

In this definition, $\mathcal{V}_\ell$ is the fixed starting view, corresponding, for instance, to the initially selected (base) marker in spatial proteomics.

Definition 2 (k-optimality)**.** Let $\sigma^\ell : \mathcal{P} \to S_{|\mathcal{V}|}$ be a selection rule. For each $p \in \mathcal{P}$ and $k \leq |\mathcal{V}|$, define the accuracy after seeing k views under the ordering given by σ^ℓ as $\text{Acc}_k(\sigma^\ell) = \mathbb{E}_{p\sim\mathcal{P}}\left[\mathbb{E}_{y\sim\mathcal{H}_k^\ell}\left[\mathbb{I}\{y = y_p\}\right]\right]$, where $\mathcal{H}_k^\ell = (h \circ F)\left(\{V_{\sigma^\ell(p)(1)}, \ldots, V_{\sigma^\ell(p)(k)}\}\right)$. Then σ^ℓ is called k-optimal if for any other selection rule ρ^ℓ, $\text{Acc}_k(\sigma^\ell) \geq \text{Acc}_k(\rho^\ell)$. Further, σ^ℓ is called an optimal selection rule if it is k-optimal for all $k \leq |\mathcal{V}|$.

We first note that due to the permutation invariance of the FM F, every selection rule is $|\mathcal{V}|$-optimal. Our goal is therefore to find a selection rule that is ideally optimal or k-optimal for values $k \ll |\mathcal{V}|$. Whether an optimal selection rule exists depends on the target task and model. Lemma 3 shows that an optimal selection rule—if existing—necessarily satisfies a greedy selection procedure.

Lemma 3. Let σ^ℓ be an optimal selection rule. Then, for every patient p and every $k \leq |\mathcal{V}|$,

$$\sigma^\ell(p)(k) = \underset{i\in[|\mathcal{V}|]\setminus\bigcup_{j=1}^{k-1}\{\sigma^\ell(p)(j)\}}{\text{argmax}} (h \circ F)(\{V_{\sigma^\ell(p)(1)}, ..., V_{\sigma^\ell(p)(k-1)}, V_i\})(y_p).$$

Our work makes the assumption that an optimal selection rule σ^ℓ is in general non-trivial, i.e., there exist patients p_1, p_2 such that $\sigma^\ell(p_1) \neq \sigma^\ell(p_2)$. This assumption reflects the intuition that as different patients exhibit varying marker profiles and diagnostic complexities, a fixed global ordering of marker acquisitions is unlikely to be optimal for all cases.

These observations suggest that one needs to learn to (i) generate personalized sequences per patient, and (ii) adaptively refine the selection strategy after incorporating feedback from h. Further, for brevity we drop the patient (p) subscript, and we denote $\sigma^\ell(p)$ as σ^ℓ. We model view selection as a RL problem with the state space as $\mathcal{S} = \mathbb{R}^d \times \mathbb{R}^{|\mathcal{Y}|} \times \{0,1\}^{|\mathcal{V}|}$, and the initial action space $\mathcal{A}_1 = \mathcal{V} \setminus \{V_\ell\}$. After t selection steps, we denote the sequence of selected views as σ_t^ℓ, with the k^{th} selection denoted as $\sigma_t^\ell(k)$. Each state is the tuple $s_t = (F(\sigma_t^\ell), \text{logit}(h \circ F(\sigma_t^\ell)), \alpha_t)$, where $\alpha_t \in \{0,1\}^{|\mathcal{V}|}$ is the mask over the available actions, where $\alpha_t[i] = 1$ if view i has not been acquired. Following Huang & Ontañón (2020), at selection step t, we disallow selection of previously chosen (invalid) views, and hence our action space evolves as $\mathcal{A}_t = \mathcal{A}_{t-1} \setminus \{\sigma_t^\ell(t)\}$. Thus, if the policy network produces unmasked logits $z(s,a)$ at state s for each action a, invalid action masking can be formulated as writing $\pi_\theta(a|s) = \texttt{SoftMax}_a(z(s,a) + \log\alpha(s,a))$, where $\alpha(s,a) = 0$ if action a has already been chosen, else 1. Following the insight of Lemma 3 that an optimal selection rule

behaves greedily, we design the instantaneous reward for selecting view i at time step t as

$$c_t(s_t, i) = \beta \left(\log \left[(h \circ F)(\{V_{\sigma_t^\ell(1)}, ..., V_{\sigma_t^\ell(t)}, V_i\})(y) \right] - \log \left[(h \circ F)(\{V_{\sigma_t^\ell(1)}, ..., V_{\sigma_t^\ell(t)}\})(y) \right] \right),$$

where β is a hyperparameter. This provides a proxy for the decision criterion, guiding the RL agent toward selecting optimal views, as the optimal view maximizes the above gain according to Lemma 3. In addition, we incorporate an episodic reward at the end of T-step sequence selection as $c_T = \delta(2\mathbb{I}((h \circ F)(\{V_{\sigma_T(1)}, ..., V_{\sigma_T(T)}\})\} = y) - 1)$, which reinforces the overall objective and rewards (penalizes) the selected sequence σ_T if the final observation has a correct (wrong) prediction, where δ is large. The combination of instantaneous rewards optimize for both: (i) immediate diagnostic improvement by prioritizing the most informative view at each step, and (ii) maximizing overall classification accuracy by learning an acquisition strategy that leads to the most reliable diagnosis.

3.2 MAVIS FOR MULTI-MODALITY VIEW SELECTION

While MAVIS for uni-modal view selection can be trained using fixed-length episodes as outlined in Section 3.1, multi-modal view selection introduces an additional challenge: The agent must learn not only which views to acquire within a modality but also when to transition from one modality to another. This requires jointly optimizing the number of views selected per modality and the optimal switching point between modalities to balance diagnostic accuracy and acquisition cost. In this section, we extend MAVIS to multi-modal view selection, focusing on two modalities: m_1, m_2, with $\mathcal{V}_1 = \{V_{1,1}, \ldots, V_{1,n_1}\}$ and $\mathcal{V}_2 = \{V_{2,1}, \ldots, V_{2,n_2}\}$, and the corresponding FMs as F_1, F_2 and classifiers as h_1, h_2. However, the framework is naturally scalable to an arbitrary number of modalities (n) making it adaptable to more complex diagnostic pipelines where decisions must be made across multiple complementary measurement techniques.

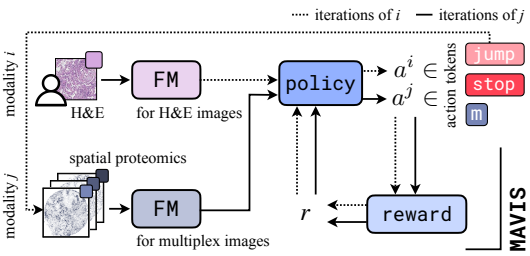

Figure 2: Overview of MAVIS's multi-modality view selection mechanism. Foundation models process H&E and multiplex images separately, with the policy network determining optimal test sequences and a reward function guiding the learning process.

In general, we assume that the views of m_2 have a higher acquisition cost that those of m_1 but also lead to higher downstream accuracy. This introduces an additional hierarchical decision-making component, where the agent must weigh the benefits of transitioning to m_2 against its increased cost. The policy must learn to prioritize lower-cost views in m_1 when sufficient for accurate diagnosis while strategically selecting high-cost views from m_2 only when necessary. Mimicking real-life clinical scenario, we begin our sequential measurements with one of the views of m_1 already present, say $\ell_1 = V_{1,\ell}$. We modify the state space to be $\mathcal{S} = \mathbb{R}^{d_{m_1}} \times \mathbb{R}^{|\mathcal{Y}|} \times \mathbb{R}^{d_{m_2}} \times \mathbb{R}^{|\mathcal{Y}|} \times \{0,1\}^{n_1+n_2+2}$, and the action space $\mathcal{A}_1 = \mathcal{V}_1 \setminus \{V_{1,\ell}\} \cup \mathcal{V}_2 \cup \{\texttt{stop}, \texttt{jump}\}$. The `stop` token signifies that we should stop the measurements overall, and the `jump` token signifies that we should start measuring m_2. After t selection steps, we denote the sequence of selected views for m_1 as $\sigma_t^{\ell_1}$, and the sequence for m_2 as ρ_t. Note that we allow free jumps from m_1 to any view of m_2, and hence the selection rules for m_2 are not conditioned on a fixed starting view. Each state is represented as the tuple $s_t = (F_1(\sigma_t^{\ell_1}), \text{logit}(h_1 \circ F_1(\sigma_t^{\ell_1})), F_2(\rho_t), \text{logit}(h_2 \circ F_2(\rho_t)), \alpha_t)$ where $\alpha_t \in \{0,1\}^{n_1+n_2+2}$ denotes a mask over the available actions, and $\alpha_t[i] = 1$ if the i^{th} action is valid and not taken previously. Since we do not have access to ρ_t before the `jump` action has been taken, we set the corresponding state vectors to $\overrightarrow{0}$. When the agent is in process of selecting views from m_1, we mask out all actions from m_2 along with previously selected views from m_1. Similarly, when the agent is selecting views from m_2, we mask out all previously selected views from m_1 and m_2 along with the `jump` token.

We want to encourage the RL agent to stop when it believes the prediction from m_1 could be right, and only move forward to m_2 when it is not confident. To do so, we design the following reward function with the following case structure (where c is the reward):

1. k^{th} marker chosen ($k > 1$) for $m_\bullet \in \{m_1, m_2\}$, and $\eta \in \{\sigma^{\ell_1}, \rho\}$ respectively:
$c = \beta \left(\log \left[(h_\bullet \circ F_\bullet)(\{V_{\eta_t(1)}, ..., V_{\eta_t(k-1)}, V_i\})(y)\right] - \log \left[(h_\bullet \circ F_\bullet)(\{V_{\eta_t(1)}, ..., V_{\eta_t(k-1)}\})(y)\right]\right),$
2. `stop` action taken after choosing k views from $\mathcal{V}_1$:
$$c = \zeta\delta(2\mathbb{I}(\text{argmax}(h_1 \circ F_1)(\{V_{\sigma^{\ell_1}_k(1)}, ..., V_{\sigma^{\ell_1}_k(k)}\}) = y) - 1)$$
3. `jump` action taken, and view ℓ_2 acquired for m_2: we have the starting view as ℓ_2 and
$$c = \beta \left(\log \left[(h_2 \circ F_2)(\{V_{\rho^{\ell_2}_t(1)}\})(y)\right] - \max_{y' \in \mathcal{Y}, y' \neq y} \log \left[(h_2 \circ F_2)(\{V_{\rho^{\ell_2}_t(1)}\})(y')\right]\right),$$
4. `stop` action taken after choosing k views from $\mathcal{V}_2$:
$$c = \delta\mathbb{I}\left(\text{argmax}(h_2 \circ F_2)(\{V_{\rho^{\ell_2}_t(1)}, ..., V_{\rho^{\ell_2}_t(k)}\}) = y\right)$$

where β, ζ, δ are hyperparameters. To achieve stable and efficient learning, we train the RL agent using PPO (Schulman et al., 2017), incorporating the invalid action masking mechanism described earlier to prevent redundant acquisitions.

3.3 MAVIS FOR H&E AND SPATIAL PROTEOMICS

In this work, we study uni-modal view selection in context of multiplex imaging experiments, where each view is a markers available for staining. Typically, high-dimensional multiplex imaging allows the iterative measurement of few (Lewis et al., 2021) to near a hundred (Black et al., 2021) marker measurements. Multiplex experiments typically begin with a cell or nucleus marker to provide structural context, and we simulate this setup by initiating each selection run with the Histone H3 marker (*i.e.*, V_ℓ = Histone H3), which stains the cell nucleus. This ensures that all subsequent marker acquisitions are conditioned on a common structural reference, aligning our experimental setup with real-world staining protocols. We utilize VirTues (Wenckstern et al., 2025) as the FM for spatial proteomics, that has shown remarkable zero-shot capabilities along with the ability to aggregate marker signals into a unified representation and denote it as F_{SP}.

For multi-modal view selection we consider two modalities, $\{\text{H\&E, SP}\}$, with $\mathcal{V}_{\text{H\&E}} = \{V_{\text{H\&E,1}}\}$, representing a single H&E stain, and $\mathcal{V}_{\text{SP}} = \{V_{\text{SP},1}, \ldots, V_{\text{SP},n}\}$, representing the available markers. For H&E, we utilize UNI2 (Chen et al., 2024), and for SP, we leverage VirTues (Wenckstern et al., 2025) as before. Mimicking real-life clinicial scenario, we begin our sequential measurements with the H&E measurement already present. Since H&E consists of a single view, the modality selection task becomes immediate, as the agent must decide whether to proceed to spatial proteomics right after observing the H&E representation.

4 EMPIRICAL RESULTS

We conduct experiments on MAVIS to showcase the effectiveness of our method across several clinical datasets, under both view and modality selection for spatial proteomics and H&E staining.

Datasets. We experiment with three datasets: (i) Cords et al. (2023) containing imaging mass cytometry (IMC) samples from non-small lung cancer patients, (ii) Danenberg et al. (2022) containing IMC samples of breast cancer patients, and (iii) Rigamonti et al. (2024), a multi-modal dataset consisting of H&E along with IMC samples from non-small lung cancer patients. For details, see Appendix B. Although IMC involves parallel staining, we model it as a sequential acquisition task as the principle of selecting and measuring specific markers remains consistent with a sequential decision-making framework.

Baselines. For view selection, we compare our approach against seven methods: (i) Random, which selects views uniformly at random; (ii-iii) Min/Max-Entropy, variants of greedy selection that use VirTues inpainting capabilities to estimate missing marker values and prioritize acquisitions based on information gain; (iv-v) Min/Max-ESM which selects views based on the distance of markers in the ESM2 embedding space (Lin et al., 2023), (vi) Iterative Panel Selection (Sims et al., 2024), a method for sequential marker selection which sets a global acquisition order based on the correlations of reconstruction post inpainting; and (vii) an expert-defined order, where a pathologist determines a fixed selection sequence as a clinically-informed baseline. For details, see Appendix C.

Evaluation. The datasets were split into an 80:20 train-test patient-wise split, with the training set further divided into an 80:20 train-validation split. To ensure well-calibrated confidence estimates, all classifiers were calibrated using temperature scaling (Guo et al., 2017). In case of uni-modal view

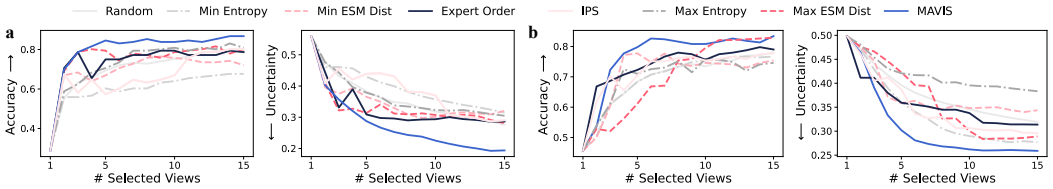

Figure 3: Performance comparison of marker selection strategies for the first 15 markers, showing accuracy (left) and uncertainty (right) trajectories for **a.** ER Status prediction (Danenberg et al., 2022) and **b.** cancer type prediction (Cords et al., 2023).

Table 1: Performance in terms of accuracy and uncertainty of MAVIS with seven baselines after selection of $B = 5$ and $B = 10$ markers. **Bold** denote the best, underlined second-best performers.

Method	Cords et al. (2023) Cancer Type				Danenberg et al. (2022) ER Status				Danenberg et al. (2022) Cancer Grade			
	$B = 5$		$B = 10$		$B = 5$		$B = 10$		$B = 5$		$B = 10$	
	Acc. ↑	Unc. ↓	Acc. ↑	Unc. ↓	Acc. ↑	Unc. ↓	Acc. ↑	Unc. ↓	Acc. ↑	Unc. ↓	Acc. ↑	Unc. ↓
ESM (Max)	0.611	0.423	0.795	0.300	0.794	0.314	0.787	0.307	0.585	0.561	0.607	0.516
ESM (Min)	0.777	0.358	0.746	0.348	0.669	0.365	0.757	0.301	0.607	0.513	0.622	0.522
Random	0.684	0.396	0.735	0.346	0.694	0.383	0.752	0.318	0.514	0.580	0.585	0.529
Entropy (Min)	0.684	0.360	0.754	0.290	0.603	0.430	0.640	0.357	0.511	0.565	0.593	0.480
Entropy (Max)	0.710	0.421	0.759	0.396	0.706	0.379	0.809	0.323	0.496	0.619	0.526	0.582
IPS	0.705	0.387	0.754	0.307	0.574	0.439	0.669	0.370	0.504	0.568	0.630	0.503
Expert Order	0.723	0.359	0.756	0.338	0.750	0.309	0.794	0.294	0.519	0.562	0.607	0.498
MAVIS	**0.798**	**0.302**	**0.808**	**0.262**	**0.831**	**0.267**	**0.846**	**0.215**	**0.622**	**0.479**	**0.674**	**0.443**

selection, we compare the performance of MAVIS with the baselines on the test set on (1) classification accuracy (↑) after sequential marker selection, and (2) prediction uncertainty (↓), quantified using $1 - \text{TCP}$, where true class probability (TCP) (Corbière et al., 2019) measures the model's confidence in the correct class. For multi-modal view selection, we measure the accuracies for cases where the selection policy opted to stop at H&E and where it proceeded to acquire additional IMC measurements, and the overall experimental cost of H&E and IMC acquisitions, assessing the trade-off between diagnostic accuracy and resource efficiency.

Results on uni-modal view selection. We begin by evaluating the performance of MAVIS against seven baselines in terms of classification accuracy and uncertainty reduction. Table 3 summarizes the results of selection under an acquisition budget constraint of measurement of $B = 5$ and $B = 10$ markers. The key observations are: (1) MAVIS consistently outperforms all baselines across three distinct diagnostic tasks on Cords et al. (2023) and Danenberg et al. (2022); (2) Although random selection is often surprisingly effective in active learning, it performs poorly in diagnostic tasks, as uniform selection of markers fails to align with the underlying biological and clinical relevance of different markers; and (3) the expert-defined ordering, ESM-distance-based selection, and IPS consistently emerge as the second or third-best performers, yet they are significantly outperformed by the adaptive ordering learned by MAVIS. This underscores the advantage of dynamically selecting markers based on patient-specific information rather than relying on a fixed global sequence. In Fig. 3, we illustrate how classification accuracy and prediction uncertainty evolve over the first 15 marker acquisitions, demonstrating that MAVIS achieves faster uncertainty reduction and higher accuracy gains compared to baseline approaches. For further results, see Appendix E, where in Fig. 7 we show that MAVIS inherently learns to select markers in a diverse manner throughout the UMAP space, in Fig. 8 we show the order of markers chosen by MAVIS for Cords et al. (2023), and in Fig. 9 we quantify the consistence of marker sequences selected by MAVIS across patients for Danenberg et al. (2022).

Results on multi-modal view selection. We evaluate the performance of MAVIS in the multi-modal setting, where the model determines whether to proceed with multiplex imaging acquisition or stop at H&E to predict cancer type of non-small lung cancer (Rigamonti et al., 2024). Unlike single-modality selection, this task requires balancing diagnostic accuracy with acquisition cost by dynamically deciding when additional information is necessary. Since random selection lacks a well-defined stopping criterion, we standardize the random baseline by fixing the selection to five markers. In Fig. 10b, we present the variation of episodic reward (blue) and episode length (black) over the course of training of MAVIS. The episodic reward steadily increases, indicating that the agent progressively learns a more optimal selection strategy. Meanwhile, the episode length stabi-

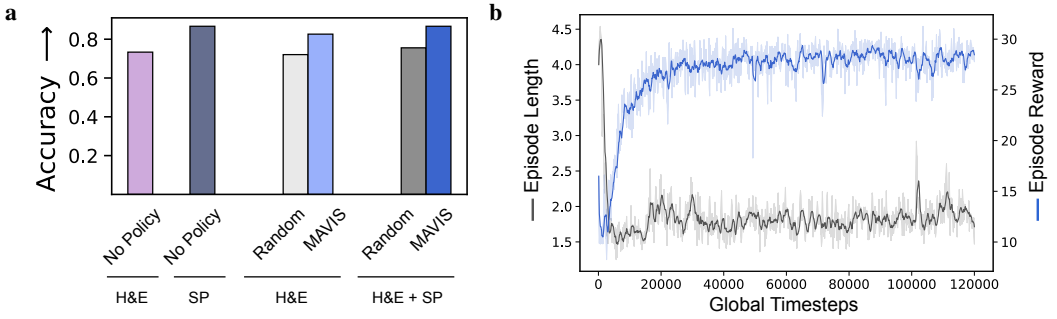

Figure 4: Results on a multi-modal view selection task between H&E and spatial proteomics-based diagnostics. **a.** Accuracy after different diagnostic test strategies. **b.** Episodic reward and episode length throughout training of MAVIS.

lizes, reflecting converged decision-making behavior. An episode length of 1 corresponds to cases where the agent selects only H&E and chooses to stop, whereas longer episodes indicate that the agent proceeds to spatial proteomics for additional biomarker acquisitions.

In Fig. 4a, we compare the accuracies of the multi-modal selection decisions made by MAVIS and random. The H&E (no policy) baseline represents the accuracy when only H&E is measured for all patients, while SP (no policy) represents the accuracy when only 27-plex imaging is used for all patients. We observe that while SP achieves significantly higher accuracy compared to H&E, this improvement comes at a substantial cost increase—rising from \$5.00 per patient (H&E) to \$2250 per patient (SP), as detailed in Table 2. The H&E (random) baseline represents the accuracy on patients where the random selection policy opted to stop at H&E stains, while H&E (MAVIS) corresponds to the accuracy on patients for whom MAVIS actively decided to stop at H&E without proceeding to multiplex imaging. We observe that MAVIS effectively identifies uncertain patients, leading to a 12.69% improvement in H&E accuracy compared to the H&E (no policy) baseline. Similarly, the H&E + SP accuracy represents the final classification performance across the entire patient cohort, integrating both cases where the model stopped at H&E and those where it continued to SP, thereby reflecting the overall effectiveness of the adaptive multi-modal selection strategy. We observe that MAVIS achieves accuracy on par with SP (no policy), demonstrating that it identifies and selectively transitions only those patients who genuinely benefit from additional multiplex imaging. This indicates that MAVIS effectively mitigates unnecessary SP acquisitions, maintaining diagnostic accuracy while significantly reducing costs. In contrast, the random policy underperforms, leading to suboptimal stop decisions and inefficient SP usage. Notably, MAVIS reduces experimental costs by over 95% while achieving the same diagnostic accuracy as 27-plex spatial proteomics staining (see Table 2). Further, in Appendix C, Fig. 10a shows the markers selected by MAVIS for multiplex imaging, and we observe that MAVIS also learns to adaptively limit its selections, choosing atmost four markers. This demonstrates the potential of adaptive multi-modal selection to minimize resource-intensive assays without compromising diagnostic reliability.

Table 2: Costs of different diagnostic test strategies.

	Total Costs
H&E	5.00 \$
SP (27-plex)	2250 \$
Random	335.56 \$
MAVIS	51.03 \$

5 Conclusion

This work introduces MAVIS, a unified framework for active multi-modal view selection in clinical diagnostics that addresses the challenge of optimizing diagnostic procedures while balancing accuracy and cost. Through a novel RL approach, MAVIS unifies modality and feature selection, enabling dynamic adaptation to individual patient needs while leveraging foundation models to guide test selection. The empirical results demonstrate that MAVIS consistently outperforms baseline approaches in both classification accuracy and uncertainty reduction, while substantially reducing testing costs. As diagnostic technologies continue to evolve, frameworks like MAVIS will become increasingly important for optimizing clinical workflows and ensuring efficient use of healthcare resources. Beyond clinical diagnostics, multi-modal RL frameworks such as MAVIS that interact with clinical foundation models will set the stage for advanced reasoning systems that actively probe biological mechanisms, guide biomarker discovery, and accelerate therapeutic development.

REFERENCES

Sarah Black, Darci Phillips, John W Hickey, Julia Kennedy-Darling, Vishal G Venkataraaman, Nikolay Samusik, Yury Goltsev, Christian M Schürch, and Garry P Nolan. Codex multiplexed tissue imaging with dna-conjugated antibodies. *Nature Protocols*, 16(8):3802–3835, 2021.

Charlotte Bunne, Yusuf Roohani, Yanay Rosen, Ankit Gupta, Xikun Zhang, Marcel Roed, Theo Alexandrov, Mohammed AlQuraishi, Patricia Brennan, Daniel B. Burkhardt, Andrea Califano, Jonah Cool, Abby F. Dernburg, Kirsty Ewing, Emily B. Fox, Matthias Haury, Amy E. Herr, Eric Horvitz, Patrick D. Hsu, Viren Jain, Gregory R. Johnson, Thomas Kalil, David R. Kelley, Shana O. Kelley, Anna Kreshuk, Tim Mitchison, Stephani Otte, Jay Shendure, Nicholas J. Sofroniew, Fabian Theis, Christina V. Theodoris, Srigokul Upadhyayula, Marc Valer, Bo Wang, Eric Xing, Serena Yeung-Levy, Marinka Zitnik, Theofanis Karaletsos, Aviv Regev, Emma Lundberg, Jure Leskovec, and Stephen R. Quake. How to Build the Virtual Cell with Artificial Intelligence: Priorities and Opportunities. *arXiv Preprint arXiv:2409.11654*, 2024.

Kathryn Chaloner and Isabella Verdinelli. Bayesian Experimental Design: A Review. *Statistical Science*, pp. 273–304, 1995.

Richard J Chen, Tong Ding, Ming Y Lu, Drew FK Williamson, Guillaume Jaume, Andrew H Song, Bowen Chen, Andrew Zhang, Daniel Shao, Muhammad Shaban, et al. Towards a general-purpose foundation model for computational pathology. *Nature Medicine*, 30(3):850–862, 2024.

Charles Corbière, Nicolas Thome, Avner Bar-Hen, Matthieu Cord, and Patrick Pérez. Addressing Failure Prediction by Learning Model Confidence. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

Lena Cords, Sandra Tietscher, Tobias Anzeneder, Claus Langwieder, Martin Rees, Natalie de Souza, and Bernd Bodenmiller. Cancer-associated fibroblast classification in single-cell and spatial proteomics data. *Nature Communications*, 14(1):4294, 2023.

Antonio Coronato, Muddasar Naeem, Giuseppe De Pietro, and Giovanni Paragliola. Reinforcement learning for intelligent healthcare applications: A survey. *Artificial Intelligence in Medicine*, 109: 101964, 2020.

Esther Danenberg, Helen Bardwell, Vito RT Zanotelli, Elena Provenzano, Suet-Feung Chin, Oscar M Rueda, Andrew Green, Emad Rakha, Samuel Aparicio, Ian O Ellis, et al. Breast tumor microenvironment structures are associated with genomic features and clinical outcome. *Nature Genetics*, 54(5):660–669, 2022.

Natalie de Souza, Shan Zhao, and Bernd Bodenmiller. Multiplex protein imaging in tumour biology. *Nature Reviews Cancer*, 24(3):171–191, 2024.

Charlotte Giesen, Hao AO Wang, Denis Schapiro, Nevena Zivanovic, Andrea Jacobs, Bodo Hattendorf, Peter J Schüffler, Daniel Grolimund, Joachim M Buhmann, Simone Brandt, et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nature Methods*, 11(4):417–422, 2014.

Dennis Gong, Jeanna M Arbesfeld-Qiu, Ella Perrault, Jung Woo Bae, and William L Hwang. Spatial oncology: Translating contextual biology to the clinic. *Cancer Cell*, 42(10):1653–1675, 2024.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On Calibration of Modern Neural Networks. In *International Conference on Machine Learning (ICML)*, pp. 1321–1330, 2017.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. *arXiv preprint arXiv:2501.12948*, 2025.

Sanket Rajan Gupte, Josiah Aklilu, Jeffrey J Nirschl, and Serena Yeung-Levy. Revisiting Active Learning in the Era of Vision Foundation Models. *arXiv preprint arXiv:2401.14555*, 2024.

Shengyi Huang and Santiago Ontañón. A Closer Look at Invalid Action Masking in Policy Gradient Algorithms. *arXiv preprint arXiv:2006.14171*, 2020.

Jonas Hübotter, Sascha Bongni, Ido Hakimi, and Andreas Krause. Efficiently learning at test-time: Active fine-tuning of llms. *arXiv preprint arXiv:2410.08020*, 2024.

Diederik P Kingma and Max Welling. Auto-Encoding Variational Bayes. *arXiv preprint arXiv:1312.6114*, 2013.

B Ravi Kiran, Ibrahim Sobh, Victor Talpaert, Patrick Mannion, Ahmad A Al Sallab, Senthil Yogamani, and Patrick Pérez. Deep Reinforcement Learning for Autonomous Driving. *IEEE Transactions on Intelligent Transportation Systems*, 23(6):4909–4926, 2021.

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. ISSN 2835-8856.

Aviral Kumar, Vincent Zhuang, Rishabh Agarwal, Yi Su, John D Co-Reyes, Avi Singh, Kate Baumli, Shariq Iqbal, Colton Bishop, Rebecca Roelofs, et al. Training Language Models to Self-Correct via Reinforcement Learning. *arXiv preprint arXiv:2409.12917*, 2024.

Sabrina M Lewis, Marie-Liesse Asselin-Labat, Quan Nguyen, Jean Berthelet, Xiao Tan, Verena C Wimmer, Delphine Merino, Kelly L Rogers, and Shalin H Naik. Spatial omics and multiplexed imaging to explore cancer biology. *Nature methods*, 18(9):997–1012, 2021.

Yang Li and Junier Oliva. Towards Cost Sensitive Decision Making. *arXiv preprint arXiv:2410.03892*, 2024.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

Charles X Ling, Qiang Yang, Jianning Wang, and Shichao Zhang. Decision Trees with Minimal Costs. In *International Conference on Machine Learning (ICML)*, pp. 69, 2004.

Yuan Ling, Sadid A Hasan, Vivek Datla, Ashequl Qadir, Kathy Lee, Joey Liu, and Oladimeji Farri. Learning to Diagnose: Assimilating Clinical Narratives using Deep Reinforcement Learning. In *International Joint Conference on Natural Language Processing (IJCNLP)*, pp. 895–905, 2017.

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 *International Conference on Machine Learning (ICML)*, 2019.

Thierry M. Nordmann, Andreas Mund, and Matthias Mann. A new understanding of tissue biology from MS-based proteomics at single-cell resolution. *Nature Methods*, 21:2220–2222, 2024.

Yu-Shao Peng, Kai-Fu Tang, Hsuan-Tien Lin, and Edward Chang. REFUEL: Exploring Sparse Features in Deep Reinforcement Learning for Fast Disease Diagnosis. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 31, 2018.

Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. Stable-Baselines3: Reliable Reinforcement Learning Implementations. *Journal of Machine Learning Research*, 22(268):1–8, 2021.

Alessandra Rigamonti, Marika Viatore, Rebecca Polidori, Daoud Rahal, Marco Erreni, Maria Rita Fumagalli, Damiano Zanini, Andrea Doni, Anna Rita Putignano, Paola Bossi, et al. Integrating AI-Powered Digital Pathology and Imaging Mass Cytometry Identifies Key Classifiers of Tumor Cells, Stroma, and Immune Cells in Non–Small Cell Lung Cancer. *Cancer Research*, 84(7): 1165–1177, 2024.

Thomas Rückstieß, Christian Osendorfer, and Patrick Van Der Smagt. Sequential Feature Selection for Classification. In *AI 2011: Advances in Artificial Intelligence*, pp. 132–141. Springer, 2011.

John Schulman, Philipp Moritz, Sergey Levine, Michael Jordan, and Pieter Abbeel. High-Dimensional Continuous Control Using Generalized Advantage Estimation. *arXiv preprint arXiv:1506.02438*, 2015.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Burr Settles. Active Learning Literature Survey. *Technical Resport*, 2009.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Y Wu, et al. DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models. *arXiv preprint arXiv:2402.03300*, 2024.

Hajin Shim, Sung Ju Hwang, and Eunho Yang. Joint Active Feature Acquisition and Classification with Variable-Size Set Encoding. *Advances in Neural Information Processing Systems (NeurIPS)*, 31, 2018.

Zachary Sims, Gordon B Mills, and Young Hwan Chang. MIM-CyCIF: masked imaging modeling for enhancing cyclic immunofluorescence (CyCIF) with panel reduction and imputation. *Communications Biology*, 7(1):409, 2024.

Yu Sun, Xiaolong Wang, Zhuang Liu, John Miller, Alexei Efros, and Moritz Hardt. Test-Time Training with Self-Supervision for Generalization under Distribution Shifts. In *International Conference on Machine Learning (ICML)*, 2020.

Richard S Sutton. Reinforcement Learning: An Introduction. *A Bradford Book*, 2018.

Hugo Touvron, Piotr Bojanowski, Mathilde Caron, Matthieu Cord, Alaaeldin El-Nouby, Edouard Grave, Gautier Izacard, Armand Joulin, Gabriel Synnaeve, Jakob Verbeek, et al. ResMLP: Feedforward networks for image classification with data-efficient training. *IEEE transactions on pattern analysis and machine intelligence*, 45(4):5314–5321, 2022.

Oriol Vinyals, Igor Babuschkin, Wojciech M Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H Choi, Richard Powell, Timo Ewalds, Petko Georgiev, et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, 2019.

Eugene Vorontsov, Alican Bozkurt, Adam Casson, George Shaikovski, Michal Zelechowski, Kristen Severson, Eric Zimmermann, James Hall, Neil Tenenholtz, Nicolo Fusi, et al. A foundation model for clinical-grade computational pathology and rare cancers detection. *Nature medicine*, 30(10): 2924–2935, 2024.

Johann Wenckstern, Eeshaan Jain, Kiril Vasilev, Matteo Pariset, Andreas Wicki, Gabriele Gut, and Charlotte Bunne. Ai-powered virtual tissues from spatial proteomics for clinical diagnostics and biomedical discovery. *arXiv preprint arXiv:2501.06039*, 2025.

Hanwen Xu, Naoto Usuyama, Jaspreet Bagga, Sheng Zhang, Rajesh Rao, Tristan Naumann, Cliff Wong, Zelalem Gero, Javier González, Yu Gu, et al. A whole-slide foundation model for digital pathology from real-world data. *Nature*, pp. 1–8, 2024.

Jinsung Yoon, James Jordon, and Mihaela Schaar. ASAC: Active Sensing using Actor-Critic models. In *Machine Learning for Healthcare Conference (MLHC)*, pp. 451–473, 2019.

APPENDIX

A BACKGROUND

Foundation models in clinical diagnostics. FMs have gained significant attention in the medical imaging and computational biology communities for their ability to learn powerful representations from diverse datasets (Chen et al., 2024; Wenckstern et al., 2025; Vorontsov et al., 2024; Xu et al., 2024). These models have shown remarkable downstream performance without additional supervision. In this work, we use UNI2 (Chen et al., 2024) as the underlying FM for H&E stained images, a model with strong zero-shot and few-shot performance on effectively classifying and segmenting histopathology images without requiring extensive task-specific training data. For spatial proteomics, we utilize VirTues (Wenckstern et al., 2025) to encode multiplexed images, since it is adept at encoding multiple marker measurements into a single embedding, capturing complex relationships within an evolving representation of a potentially growing number of selected markers.

While foundation models demonstrate impressive capabilities, their real-world deployment often faces challenges due to distribution shifts. Hübotter et al. (2024) address this through combining test-time training (Sun et al., 2020) with active learning (Settles, 2009), which allows models to adapt to test samples during inference through self-supervised learning objectives, without requiring additional labeled data. Their approach actively selects samples to reduce uncertainty but requires prior knowledge of the complete data space—a limitation in clinical settings. Gupte et al. (2024) propose an uncertainty-based active learning strategies for improving the performance of foundation models on downstream tasks, though focusing on sample selection rather than feature acquisition.

Reinforcement learning. RL has proven effective for a variety of sequential decision-making problems (Sutton, 2018), from game-playing agents (Vinyals et al., 2019) to autonomous vehicles (Kiran et al., 2021). More recently, reinforcement learning from human feedback (RLHF) has been employed to refine reasoning in large language models (LLMs), guiding them toward generating more coherent and contextually appropriate responses (Kumar et al., 2024; Guo et al., 2025; Shao et al., 2024). In the context of clinical diagnostics (Ling et al., 2017; Coronato et al., 2020; Peng et al., 2018), RL offers a principled way to iteratively decide which diagnostic analysis or features to acquire at test-time in the clinic, where fine-tuning and supervision are not possible.

Formally, the RL problem can be framed as a Markov decision process (MDP) $(\mathcal{S}, \mathcal{A}, P, c, \gamma)$, with $\mathcal{S}$ as the set of possible states (partial test results), $\mathcal{A}$ being the set of actions (which test/marker to select), $P : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ being the transition function defining how states evolve after taking actions, $c : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ as the reward function, and γ as the discount factor. The goal of RL is to learn a policy $\pi_\theta : \mathcal{S} \to \Delta(\mathcal{A})$ mapping states to a probability distribution over the discrete set $\mathcal{A}$. With τ being the sequence of $((s_t, a_t, c_t))_{t=0}^{T}$, the goal of policy optimization is to maximize the expected γ-discounted cumulative return $J(\theta) = \mathbb{E}_{\tau \sim P_\pi(\tau)} \left[\sum_{t=0}^{T} \gamma^t c(s_t, a_t)\right]$, and $\pi^* = \operatorname{argmax}_\pi J(\theta)$. Proximal policy optimization (PPO) (Schulman et al., 2017) is an effective model-free policy gradient algorithm, which learns a policy π_θ and a value function $V_\theta : \mathcal{S} \to \mathbb{R}$. After collecting a set of new episodes $\mathcal{E}$, using the policy prior to the update step, PPO updates the policy and value networks by optimizing

$$L(\theta) = \mathbb{E}_{(s_t, a_t) \sim \mathcal{E}} \left[\min\left(r_t(\theta)\hat{A}_t, \text{clip}\left(r_t(\theta), 1-\epsilon, 1+\epsilon\right)\hat{A}_t\right) - \lambda_1 \left(V_\theta(s_t) - R_t\right)^2 + \lambda_2 H(\pi_\theta)\right],$$

where $H(\pi_\theta) = -\sum_{a \in \mathcal{A}} \pi_\theta(a|s_t) \log \pi_\theta(a|s_t)$, and $\lambda_1, \lambda_2, \epsilon$ are hyperparameters, and the advantage $\hat{A}_t$ is calculated using an estimation scheme known as generalized advantage estimator (GAE) (Schulman et al., 2015).

Active feature acquisition. Active feature acquisition studies the problem of collecting features associated with a cost to maximize the prediction performance for a target variable while minimizing the cost of selected features. Ling et al. (2004) utilize decision trees with minimal costs, while Ma et al. (2019) propose a greedy mutual-information maximization procedure to select the next feature to collect, where the mutual information is approximated using surrogate features generated by a variational autoencoder (Kingma & Welling, 2013). Rückstieß et al. (2011), Shim et al. (2018), Yoon et al. (2019) formalize the problem as as MDP and subsequently employ reinforcement learning methods to learn a policy governing the feature selection. Li & Oliva (2024) model the feature acquisition and prediction process as a partially-observable MDP and propose a generative-approach

to impute missing features, which represent the dynamic belief state of the agent. Kossen et al. (2023) extend the problem of active feature acquisition by an additional temporal dimension.

B DATASETS

In our experiments, we evaluate MAVIS on three real-world datasets, each selected to capture different aspects of multi-modal diagnostic test selection. For further details on the datasets, especially dataset processing, normalization and curation, see Wenckstern et al. (2025) and Rigamonti et al. (2024).

1. **IMC dataset from Cords et al. (2023)**: This dataset consists of 2,034 IMC samples collected from 1,068 patients diagnosed with non-small cell lung cancer (NSCLC). We primarily use it for cancer type classification, distinguishing between Adenocarcinoma and Squamous Cell Carcinoma.
2. **IMC dataset from Danenberg et al. (2022)**: This dataset comprises of 677 IMC samples from breast cancer patients and is used for predicting ER status and cancer grade. The ER status classification task has two labels: positive and negative, while the cancer grade prediction task involves three categories: Grade 1, Grade 2, and Grade 3.
3. **H&E and IMC dataset from Rigamonti et al. (2024)**: This dataset integrates H&E-stained images with corresponding IMC samples from non-small cell lung cancer (NSCLC) patients, enabling a multi-modal diagnostic analysis. It consists of 158 samples, with cancer type labels classified as adenocarcinoma or squamous cell carcinoma.

For all datasets, we adopt a patient-wise split with an 80:20 ratio for training and testing. The training set is further divided into an 80:20 split to form a validation set.

C EXPERIMENTAL DETAILS

Our experimental evaluation of MAVIS includes extensive comparisons against multiple baseline methods. Below, we outline the key experimental settings and implementation details. For details on the training of VirTues and UNI2, we refer the reader to Wenckstern et al. (2025) and Chen et al. (2024), respectively.

C.1 BASELINES

To benchmark the performance of MAVIS, we compare against seven baseline methods for unimodal view selection:

- **Random**: At any time step, this selects views uniformly at random, and hence $\Pr(V_\bullet | V_{\rho(1)}, \dots V_{\rho(k)}) = \frac{1}{|\mathcal{V}|-k}$. In our experiments, we run 5 iterations of random and showcase the mean values.
- **Entropy-based selection** (min and max): VirTues (denoted as F_{SP}) allows inpainting of unseen markers, and hence at any timestep t, we can reconstruct the unseen markers based on the history of markers queried. Entropy-based selection utilizes this inpainting ability to approximate the embedding of the marker set after reconstructing the unseen marker, and further selects the markers based on the performance of the downstream classifier h_{SP}. In essence, if $(V_{\rho(1)}, \dots, V_{\rho(k)})$ is the sequence of markers queried, then, Min-Entropy selects the next marker as
$$\rho(k+1) = \underset{j \in [\![1,n]\!] \setminus \bigcup_{i=1}^{k} \{\rho(i)\}}{\operatorname{argmin}} \; H\left((h_{\text{SP}} \circ \mathcal{V}_{\text{SP}}) \left(\{V_{\rho(1)}, \dots, V_{\rho(k)}, \widetilde{V}_j\} \right) \right),$$
where $H(\cdot)$ denotes the entropy, and $\widetilde{V}_j = F_{\text{SP}}^{\text{recon}}(V_j | \{V_{\rho(1)}, \dots, V_{\rho(k)}, \widetilde{V}_j\})$ is the reconstruction from VirTues after seeing the markers from selection rule ρ. Similarly, Max-Entropy chooses the marker which maximizes the entropy.
- **ESM distance-based selection** (min and max): Biomarkers used for multiplex imaging can be encoded using ESM2 (Lin et al., 2023), and for each marker v, we can obtain a ranked list of nearest and furthest markers in the ESM2-space based on euclidean distance.

In essence, if $(V_{\rho(1)}, \ldots, V_{\rho(k)})$ is the sequence of markers queried, then Min-ESM selects the next marker as

$$\rho(k+1) = \underset{j \in [\![1,n]\!] \setminus \bigcup_{i=1}^{k} \{\rho(i)\}}{\operatorname{argmin}} \|E[V_j] - E[V_k]\|_2,$$

where $E[V_j]$ denotes the ESM2 embedding for marker V_j. Similarly, Max-ESM selects the marker which maximizes the euclidean distance.

- **Iterative panel selection**: Sims et al. (2024) proposed an iterative selection approach that selects the next marker based on the correlation of the reconstruction through inpainting on the cellular segmentation mask for the i^{th} image, IPS sets a global order on the training data, by selecting

$$\rho(k+1) = \underset{j \in [\![1,n]\!] \setminus \bigcup_{i=1}^{k} \{V_{\rho(i)}\}}{\operatorname{argmax}} r_s(F_{\text{SP}}^{\text{recon}}(\mathcal{V} \setminus \mathcal{V}_{\text{obs},j} | \mathcal{V}_{\text{obs},j}), \mathcal{V} \setminus \mathcal{V}_{\text{obs},j}),$$

where $\mathcal{V}_{\text{obs},j} = \cup_{i=1}^{k} \{V_{\rho(i)}\} \cup \{V_{\rho(j)}\}$ is the set of markers previously observed and V_j under consideration, and r_s is spearman correlation between the mean intensities of the reconstruction and ground truth. In essence, IPS selects markers based on informativeness to reconstruct other markers.

- **Expert order**: A clinically informed fixed sequence determined by a pathologist. We provide the top 15 markers selected below:
  - Breast cancer: CK5, CD8-18, HER2, ER, Cav1, Ki-67, panCK, CD68, CD8, CD4, CD16, CD11c, CD38, PDPN, SMA
  - Lung cancer : panCK, CDH6, Cadherin-11, CD73, CD279, vWF, Cav-1, Ki-67, VCAM1, FAP, LYVE-1, CD34, CD8a, CD20, CD68

For modality selection in the multi-modal setup, we include a **random modality switching baseline**, where the system switches from H&E to spatial proteomics with a fixed probability (0.5).

C.2 Evaluation

We consider two main criterion of evaluating the methods in uni-modal view selection: (1) Accuracy, and (2) Uncertainty. Accuracy is simply the fraction of correctly classified images, measuring the overall diagnostic performance, while uncertainty is defined as 1 - TCP, where TCP quantifies the model's confidence in the predicted class (Corbière et al., 2019). A lower uncertainty score indicates more confident and reliable predictions. Let σ^{ℓ} be the selection rule, F, h be the foundation model and classifier respectively, and y be the ground truth label. For the uni-modal settings, after k selections, we have

- accuracy after k: $\text{Acc}_k(\mathcal{P}_{\text{test}}) = \mathbb{E}_{p \sim \mathcal{P}_{\text{test}}} \left[\mathbb{I} \left(\operatorname{argmax}(h \circ F)(\{V_{\sigma_t^{\ell}(p)(1)}, ..., V_{\sigma_t^{\ell}(p)(k)}\}) = y \right) \right]$,
- uncertainty after k: $\text{Unc}_k(\mathcal{P}_{\text{test}}) = \mathbb{E}_{p \sim \mathcal{P}_{\text{test}}} \left[1 - (h \circ F)(\{V_{\sigma_t^{\ell}(1)}, ..., V_{\sigma_t^{\ell}(k)}\})(y) \right]$.

For multi-modal view selection, we calculate the individual modality accuracies as above, ans the overall accuracy is given as

$$\text{Acc}_{\text{overall}} = \frac{\text{Acc}^{(1)}(\mathcal{P}_{\text{test},1})|\mathcal{P}_{\text{test},1}| + \text{Acc}^{(2)}(\mathcal{P}_{\text{test},2})|\mathcal{P}_{\text{test},2}|}{|\mathcal{P}_{\text{test},1}| + |\mathcal{P}_{\text{test},2}|}.$$

where $\text{Acc}^{(k)}(\mathcal{P}_{\text{test},k})$ is the accuracy of the selection rule on modality m_k on those test samples which were selected by MAVIS or Random on m_k. Overall, this is the fraction of the total correct classifications in both modalities.

C.3 Hyperparameters and Implementation Details

Training procedure. We train MAVIS using Proximal Policy Optimization (Schulman et al., 2017). To stabilize learning, we implement invalid action masking, ensuring that already-selected biomarkers or invalid modality-switch tokens are not re-selected. We set $\beta = 40$, $\zeta = \delta = 5$. While we did not perform exhaustive hyperparameter tuning, exploring the impact of these parameters would be interesting. In Fig. 5, we show the episodic and instantaneous rewards for MAVIS in the multi-modal view selection setting. This represents the reward structure as describe in Section 3. In summary: (i) the rewards on choosing a new view for m_1 and m_2 are the difference in log probabilities of the true class, (ii) the episodic reward for stopping at m_1 if the prediction is correct is

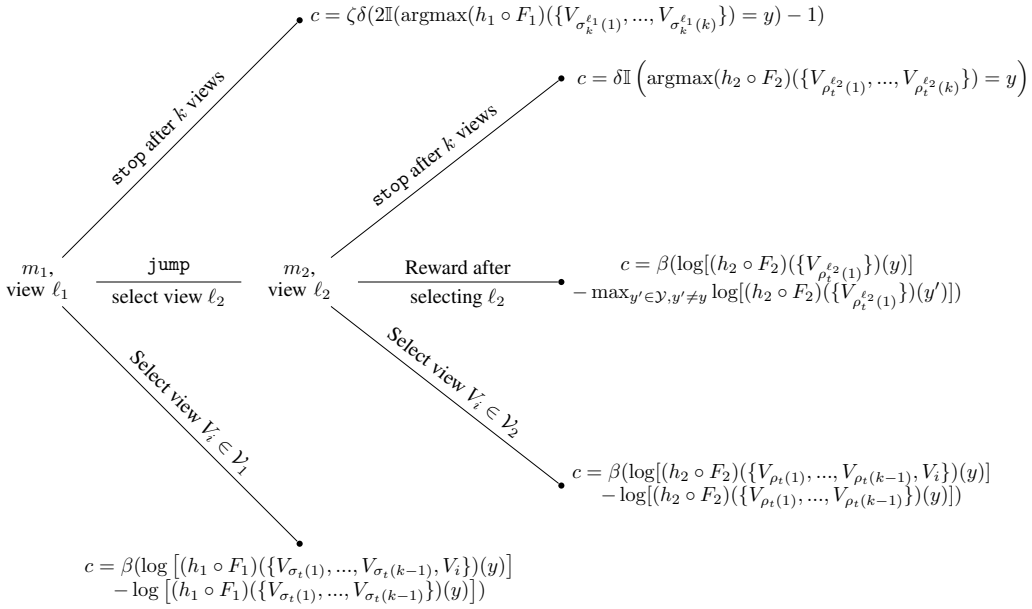

Figure 5: Reward structure for multi-modal view selection by MAVIS, as described in Section 3.

higher than m_2 (since $\zeta > 1$), which is meant to encourage stopping at m_1 when the prediction is believed to be accurate, while there is a large penalty for an incorrect stop, indicating the fact that uncertain predictions should move onto m_2 regardless. The reward on moving to ℓ_2 is the minimum difference in probabilities between the true label and other labels.

Calibration and evaluation. All classifiers, implemented as ResMLPs (Touvron et al., 2022), with 2 residual blocks and hidden dimension of 512, are calibrated using temperature scaling (Guo et al., 2017) to ensure well-calibrated confidence estimates.

Computational setup. Model training and evaluation were performed on A100 GPU clusters. We utilize Stable Baselines 3 (Raffin et al., 2021) to implement the policy and reward functions. We set $\gamma = 0.99$, with $\lambda_1 = 0.5$, $\lambda_2 = 0$ and $\epsilon = 0.2$. We compute rollout updates every 100 steps for multi-modal setting, while 1000 steps for uni-modal view selection.

D Additional Theoretical Analysis

D.1 Proof of Lemma 3

Proof. Assume, for sake of contradiction, that there exists a patient p, $1 \leq k \leq |\mathcal{V}|$ and an index $i \in [|\mathcal{V}|] \setminus \sigma(p)([k-1])$ such that

$$(h \circ F)(\{V_{\sigma(p)(1)}, ..., V_{\sigma(p)(k-1)}, V_i\})(y_p) > (h \circ F)(\{V_{\sigma(p)(1)}, ..., V_{\sigma(p)(k-1)}, V_{\sigma(p)(k)}\})(y_p)$$

We can then define a new selection rule $\tilde{\sigma}$ by the swapping k-th element of the order $\sigma(p)$ with i, i.e., setting

$$\tilde{\sigma}(q)(j) = \begin{cases} \sigma(q)(j) & \text{if } q \neq p \\ \sigma(q)(j) & \text{if } q = p \text{ and } j \notin \{k, \sigma(p)^{-1}(i)\} \\ i & \text{if } q = p \text{ and } j = k \\ \sigma(p)(k) & \text{if } q = p \text{ and } j = \sigma(p)^{-1}(i). \end{cases}$$

For the new selection rule $\tilde{\sigma}$, it holds for patients $q \neq p$ that

$$(h \circ F)(\{V_{\tilde{\sigma}(q)(1)}, ..., V_{\tilde{\sigma}(q)(k-1)}, V_{\tilde{\sigma}(q)(k)}\})(y_p) = (h \circ F)(\{V_{\sigma(q)(1)}, ..., V_{\sigma(q)(k-1)}, V_{\sigma(q)(k)}\})(y_p)$$

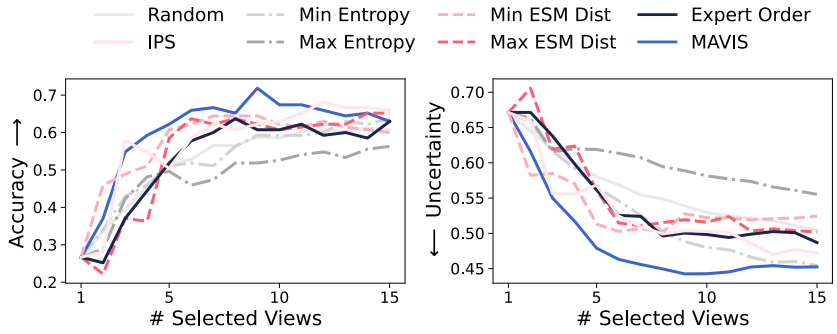

Figure 6: Performance comparison of marker selection strategies for the first 15 biomarkers, showing accuracy (left) and uncertainty (right) trajectories for cancer grade prediction in Danenberg et al. (2022).

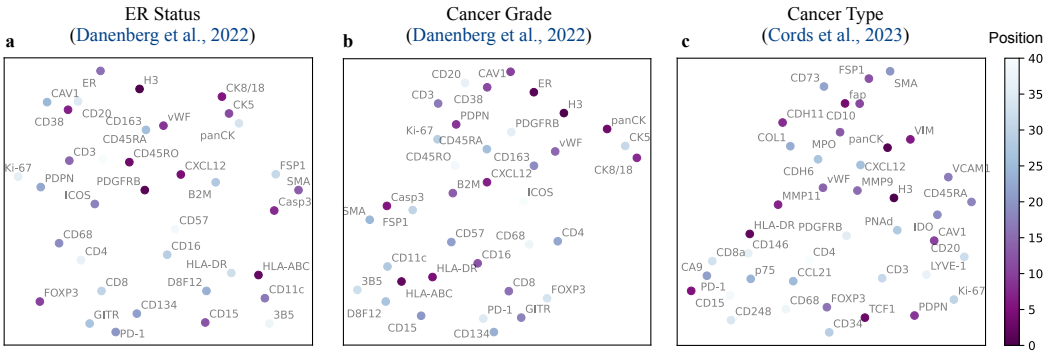

Figure 7: UMAPs of the ESM embeddings of the respective markers in Danenberg et al. (2022) and Cords et al. (2023). The color map indicates the positions in the sequence determined by MAVIS in different diagnostic classification tasks, i.e., ER status, cancer grade, and cancer type prediction.

and further for p that

$$(h \circ F)(\{V_{\tilde{\sigma}(q)(1)}, ..., V_{\tilde{\sigma}(q)(k-1)}, V_{\tilde{\sigma}(q)(k)}\})(y_p)$$
$$> (h \circ F)(\{V_{\sigma(q)(1)}, ..., V_{\sigma(q)(k-1)}, V_{\sigma(q)(k)}\})(y_p).$$

Taking the expectation, it follows that $\mathrm{Acc}_k(\tilde{\sigma}) > \mathrm{Acc}_k(\sigma)$. This contradicts our assumption that σ is optimal. □

E Additional Empirical Results

In this section, we provide further analysis of the sequential marker selection behavior exhibited by MAVIS for uni-modal view selection. In particular, we analyze the ESM embeddings and their distributions given the sequential order determined by MAVIS, the distribution of marker selection frequencies, and similarities between different patients given the marker orderings determined by MAVIS in relation to further clinical annotations that are available.

In Table 3, we showcase the accuracy and uncertainty of MAVIS as compared to other baselines on selection of 5 or 10 markers. We observe that MAVIS consistently outperforms all other methods with a significant margin.

In Fig. 6, we present the accuracy vs. uncertainty trade-off plots for cancer grade classification on Danenberg et al. (2022). The results indicate that MAVIS consistently selects more informative biomarker sequences compared to baseline methods, leading to higher classification accuracy while maintaining lower uncertainty.

Fig. 7 presents UMAP projections of the ESM embeddings for biomarkers from the Danenberg et al. (2022) and Cords et al. (2023) datasets. In these visualizations, markers are color-coded based on their selection order as determined by MAVIS across different diagnostic tasks, including ER status

Table 3: Performance in terms of accuracy and uncertainty of MAVIS with seven baselines after selection of $B = 5$ and $B = 10$ markers. **Bold** denote the best, underlined second-best performers.

Method	Cords et al. (2023) Cancer Type				Danenberg et al. (2022) ER Status				Danenberg et al. (2022) Cancer Grade			
	$B = 5$		$B = 10$		$B = 5$		$B = 10$		$B = 5$		$B = 10$	
	Acc. ↑	Unc. ↓	Acc. ↑	Unc. ↓	Acc. ↑	Unc. ↓	Acc. ↑	Unc. ↓	Acc. ↑	Unc. ↓	Acc. ↑	Unc. ↓
ESM (Max)	0.611	0.423	0.795	0.300	0.794	0.314	0.787	0.307	0.585	0.561	0.607	0.516
ESM (Min)	0.777	0.358	0.746	0.348	0.669	0.365	0.757	0.301	0.607	0.513	0.622	0.522
Random	0.684	0.396	0.735	0.346	0.694	0.383	0.752	0.318	0.514	0.580	0.585	0.529
Entropy (Min)	0.684	0.360	0.754	0.290	0.603	0.430	0.640	0.357	0.511	0.565	0.593	0.480
Entropy (Max)	0.710	0.421	0.759	0.396	0.706	0.379	0.809	0.323	0.496	0.619	0.526	0.582
IPS	0.705	0.387	0.754	0.307	0.574	0.439	0.669	0.370	0.504	0.568	0.630	0.503
Expert Order	0.723	0.359	0.756	0.338	0.750	0.309	0.794	0.294	0.519	0.562	0.607	0.498
MAVIS	**0.798**	**0.302**	**0.808**	**0.262**	**0.831**	**0.267**	**0.846**	**0.215**	**0.622**	**0.479**	**0.674**	**0.443**

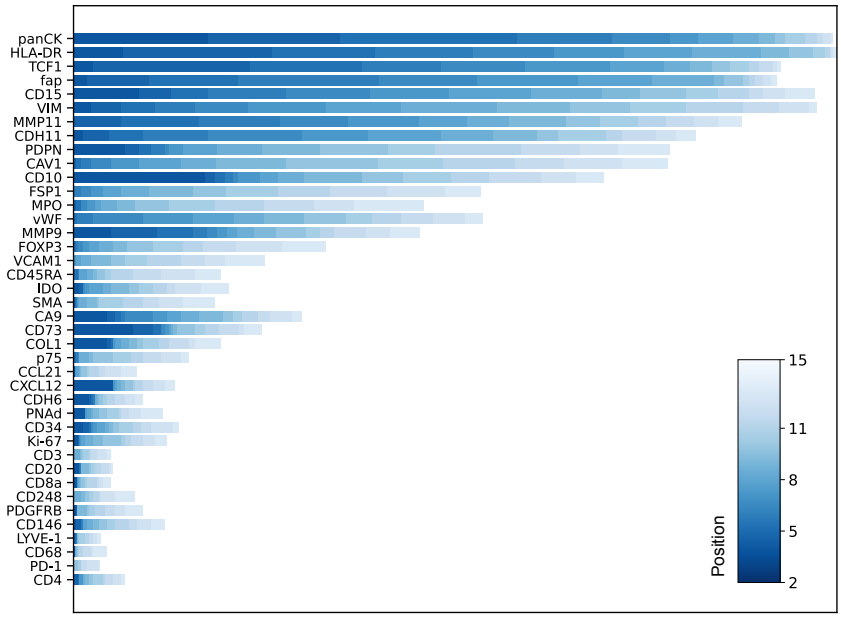

Figure 8: Distribution of marker selection frequencies across the first 15 positions in spatial proteomics panel design, highlighting the most commonly selected markers and their typical positions in the measurement sequence, here for cancer type prediction in Cords et al. (2023).

classification, cancer grade prediction, and cancer type identification. The UMAP projections reveal that MAVIS learns to select biomarkers in a structured and diverse manner, effectively navigating the embedding space to capture intrinsic marker relationships. Instead of relying on static, predefined sequences, MAVIS dynamically adapts its selection strategy to the specific diagnostic task, ensuring that informative and complementary markers are prioritized. This adaptive ordering highlight's the models ability to capture intrinsic relationships between the markers to determine an effective ordering that adapts to the specific diagnostic task.

Fig. 8 illustrates the distribution of marker selection frequencies across the first 15 positions in the spatial proteomics panel design, focusing on the cancer type prediction task in the Cords et al. (2023) dataset. The figure highlights the most frequently selected markers at each step, providing insights into the ordering preferences learned by MAVIS. Notably, several biomarkers consistently appear in early selection positions across different patients, indicating their high diagnostic value when acquired early in the sequence. In contrast, other biomarkers exhibit greater variability in their position, suggesting that MAVIS employs an adaptive selection strategy tailored to individual patient profiles rather than adhering to a rigid global ordering. This shows the ability of MAVIS to not only identify diagnostically important markers but also to optimize their acquisition order, balancing early information gain with efficient resource utilization to maximize diagnostic performance while controlling experimental costs.

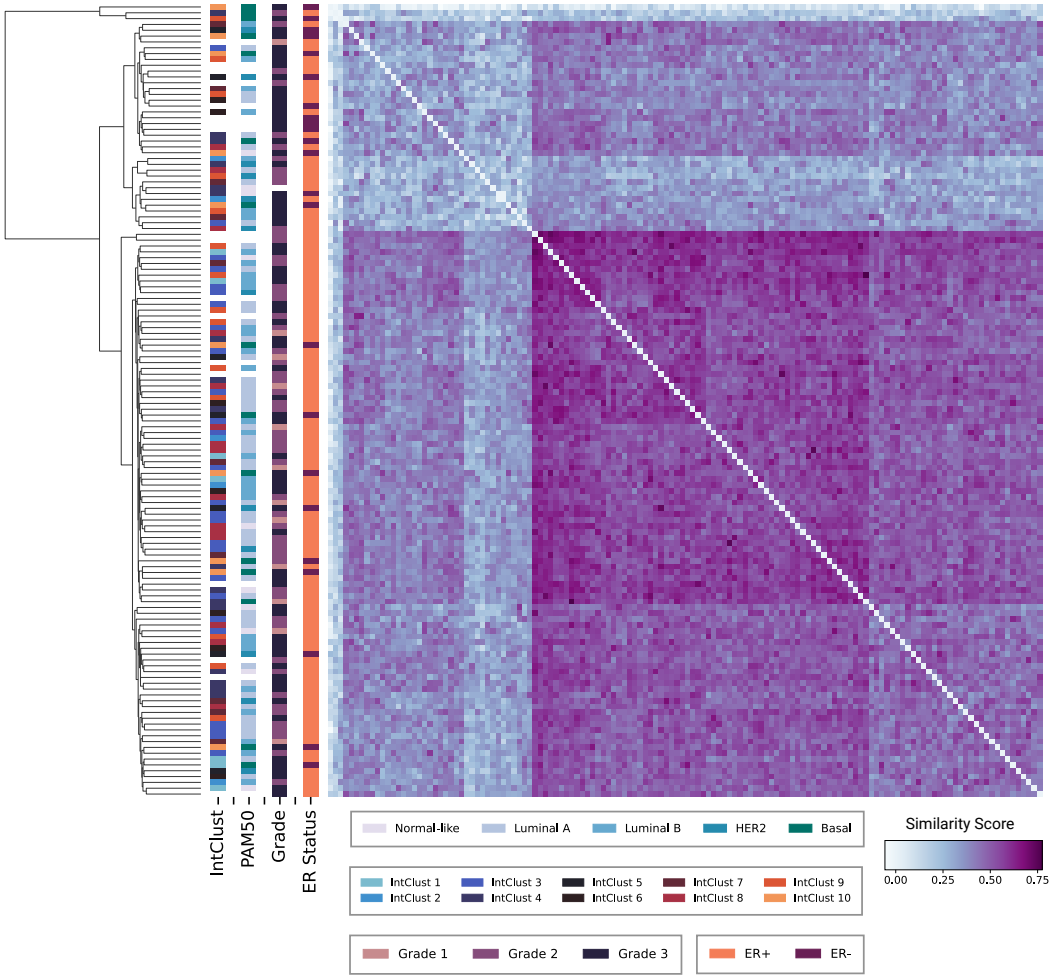

Figure 9: Clustergram of the marker sequence selected by MAVIS quantified through the Kendall rank correlation coefficient between the selection orders chosen by for all patients in Danenberg et al. (2022). The heatmap corresponds to the pairwise Kendall tau coefficients for the task of predicting the ER status. Alongside, we also display different clinical labels (integrative cluster (IntClust) subtypes, PAM50 subtyping in breast cancer, cancer grade and ER status) for the patients.

Fig. 9 offers a clustergram that quantifies the consistency of the marker sequences selected by MAVIS across patients, using the Kendall rank correlation coefficient (Kendall's tau) as the similarity measure between two sequences of chosen markers. The main heatmap corresponds to the task of predicting ER status, while additional clinical labels (such as integrative cluster subtypes, PAM50 subtyping in breast cancer, and cancer grade) are also analyzed. The clustergram reveals strong correlations in marker ordering among patients with similar diagnostic characteristics, while also highlighting distinct differences when comparing across varied clinical labels. This result confirms that MAVIS learns a patient-specific yet coherent ordering of markers, effectively balancing the need for individualized test sequences with the overall diagnostic objectives.

In Fig. 10a, we illustrate the sequential marker selection order determined by MAVIS in the multi-modal view selection setting. Unlike the uni-modal view selection setting, where the number of selected markers was fixed, MAVIS was not explicitly constrained to choose a predetermined number of markers in the multi-modal setting. Interestingly, MAVIS automatically learned to adaptively limit its selections, choosing at most four markers in some cases while opting for only a single marker in others.

In Fig. 10b, we present the variation of episodic reward (blue) and episode length (black) over the course of training of MAVIS. The episodic reward steadily increases, indicating that the agent

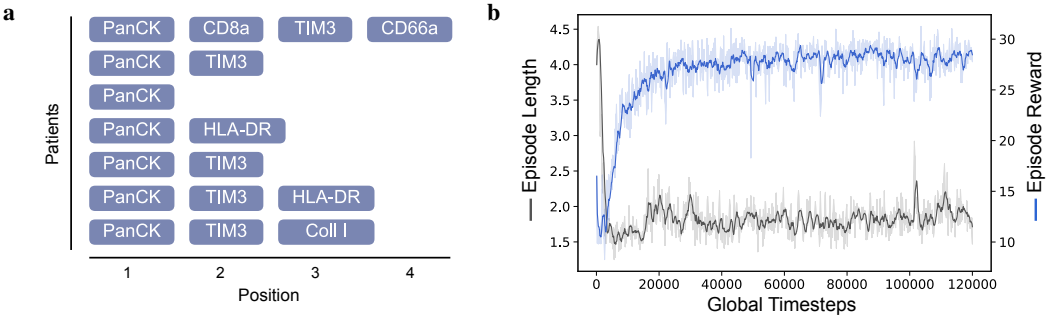

Figure 10: **a.** Selection rule by MAVIS for spatial proteomics modality for multi-modal view selection. **b.** Episodic reward and episode length throughout training of MAVIS.

progressively learns a more optimal selection strategy. Meanwhile, the episode length stabilizes, reflecting converged decision-making behavior. An episode length of 1 corresponds to cases where the agent selects only H&E and chooses to stop, whereas longer episodes indicate that the agent proceeds to spatial proteomics for additional biomarker acquisitions.