Test-Time View Selection for Multi-Modal Decision Making

Eeshaan Jain, Johann Wenckstern, Benedikt von Querfurth, Charlotte Bunne

EPFL

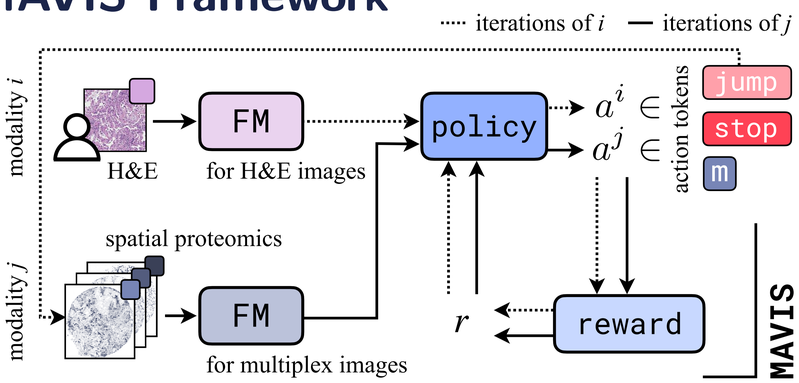
Contributions

... **unified framework for modality and feature selection**

... view selection *at inference time* without ground-truth preferences

... learning to guide **test-time selection** through **foundation models**

MAVIS Framework



Aim: learn the best selection rule σ^ℓ for any patient starting from any view ℓ

Challenges: inferring the optimal selection rule is $\Omega(2^{|\mathcal{V}|})$
optimal selection rules different for every patient

Metrics: $\mathrm{Acc}_k = \mathbb{E}_{p\sim\mathcal{P}}\left[\mathbb{E}_{y\sim\mathcal{H}_k^\ell}\left[\mathbb{I}\{y = y_p\}\right]\right]$ foundation model

$(h \circ F)\left(\{V_{\sigma^\ell(p)(1)}, \ldots, V_{\sigma^\ell(p)(k)}\}\right)$

$\mathrm{Unc}_k = 1 - \mathrm{TCP}_k$

σ^ℓ is k-optimal: If $\forall$ selection rules ρ^ℓ: $\mathrm{Acc}_k(\sigma^\ell) \geq \mathrm{Acc}_k(\rho^\ell)$

σ^ℓ is optimal: If σ^ℓ is k-optimal for every $k \leq |\mathcal{V}|$

↳ If exists, is necessarily greedy!

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

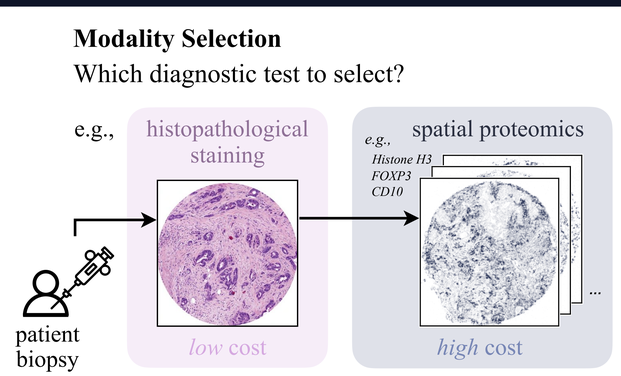
Learning k-optimal selection rules for $k \ll |\mathcal{V}|$:

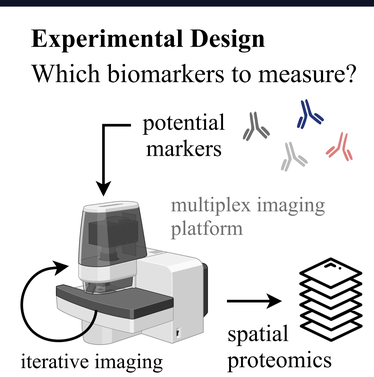
$$c_t(s_t, i) = \beta\left(\log\left[(h \circ F)(\{V_{\sigma_t^\ell(1)}, ..., V_{\sigma_t^\ell(t)}, V_i\})(y)\right] - \log\left[(h \circ F)(\{V_{\sigma_t^\ell(1)}, ..., V_{\sigma_t^\ell(t)}\})(y)\right]\right),$$

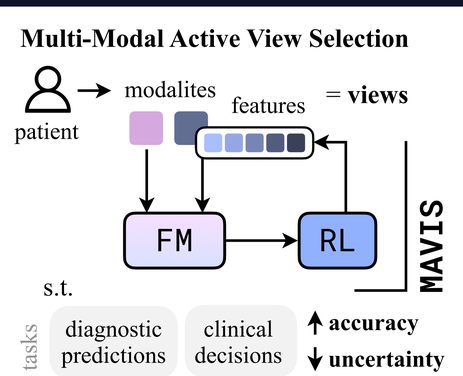
← instantaneous reward

episodic reward ↓

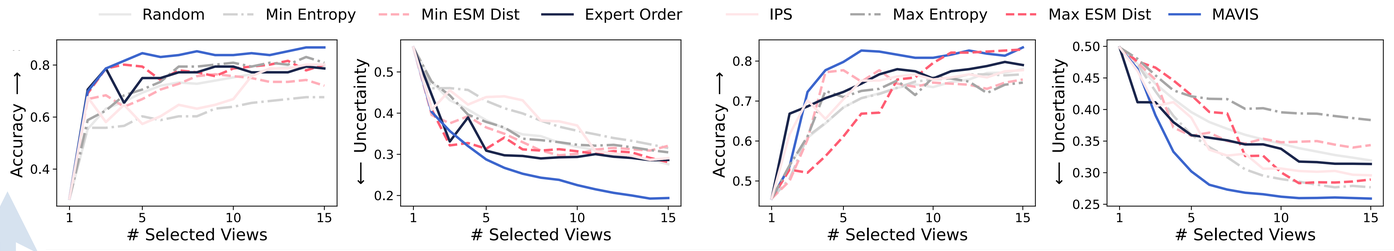
$$c_T = \delta(2\mathbb{I}((h \circ F)(\{V_{\sigma_T(1)}, ..., V_{\sigma_T(T)}\})\} = y) - 1)$$





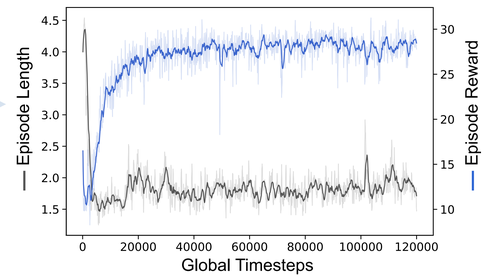


Results: MAVIS learns to select better selection rules



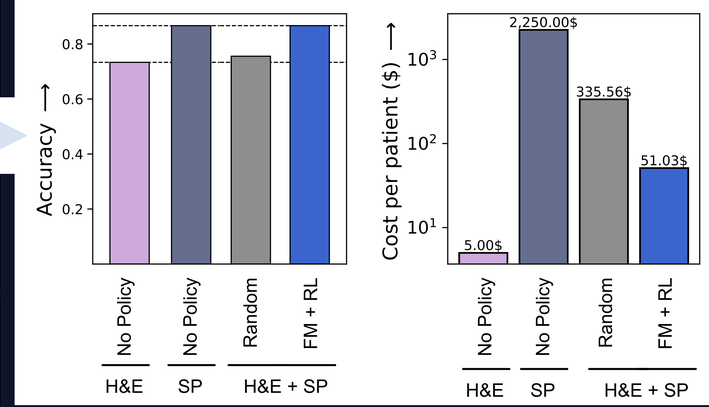
Over 15 marker acquisitions, **MAVIS** achieves faster uncertainty reduction and higher accuracy gains, outperforming global and expert orders.

MAVIS progressively learns better selection rules while learning when to switch to multiplex imaging over H&E.



MAVIS achieves accuracy on par with SP, transitioning only those patients from H&E who genuinely benefit from additional multiplex imaging.

MAVIS reduces experimental costs by over 95% and learns to adaptively limit biomarker selection.



switch reward → $c = \zeta\delta(2\mathbb{I}(\arg\max(h \circ F)(\{V_{\sigma_k^\ell(1)}, ..., V_{\sigma_k^\ell(k)}\}) = y) - 1)$

To be presented at **MLGenX (Spotlight) & GemBio, ICLR 2025**

... we introduce **MAVIS** a unified framework for active multi-modal view selection in clinical diagnostics enabling dynamic adaptation to individual patient needs while leveraging FMs to guide test selection.

Acknowledgements: Yexiang Cheng (EPFL), Phil Cheng (HUG)