SAGEAgent: A Self-Evolving Agent for Cost-Aware Modality Acquisition in Multimodal Survival Prediction

Chongyu Qu[1], Can Cui[1], Zhengyi Lu[1], Junchao Zhu[1], Tianyuan Yao[1], Junlin Guo[1], Juming Xiong[1], Yanfan Zhu[1], Yuechen Yang[1], Bennett A. Landman[1,2], and Yuankai Huo[1]✉

[1] Vanderbilt University, Nashville TN 37235, USA
[2] Vanderbilt University Medical Center, Nashville TN, 37232 USA
yuankai.huo@vanderbilt.edu

Abstract. Does every cancer patient truly need a complete diagnostic workup for accurate survival prediction? In multimodal clinical oncology, diagnostic modalities follow a clinically mandated order of escalating burden—from demographics collected at intake to genomic profiling requiring specialized tissue analysis. Current multimodal survival methods either assume all modalities are available or passively handle missing data, but none actively reason about whether acquiring the next modality is justified for a given patient along this ordered workflow. We formulate this as a sequential decision problem and propose SAGEAgent (**S**equential **A**cquisition **G**uided by **E**xperience), a self-evolving LLM-based clinical agent that decides which diagnostic modalities to acquire for each patient, balancing predictive accuracy against clinical invasiveness. SAGEAgent reasons about each patient's evolving diagnostic state through clinical tools that translate numerical predictions into text, an episodic memory that retrieves similar past cases, and a semantic memory that accumulates reusable decision patterns from experience. Experiments on a glioma cohort combining TCGA-LGG, TCGA-GBM, and BraTS with four diagnostic modalities demonstrate that SAGEAgent achieves competitive survival prediction accuracy while reducing average acquisition burden by 55%. Code is publicly available at: https://github.com/Chongyu1117/SAGEAgent

Keywords: Modality Acquisition · LLM Agent · Survival Prediction.

1 Introduction

Accurate survival prediction in oncology increasingly relies on integrating multiple diagnostic modalities [3,15], ranging from routine demographics and radiology to invasive biopsy and molecular profiling [16]. These modalities follow a strict clinical order: imaging precedes biopsy, which in turn provides the specimen for genomic analysis. Previous multimodal fusion methods [5,4,8] combine all available modalities through a fixed strategy or handle missing data at test

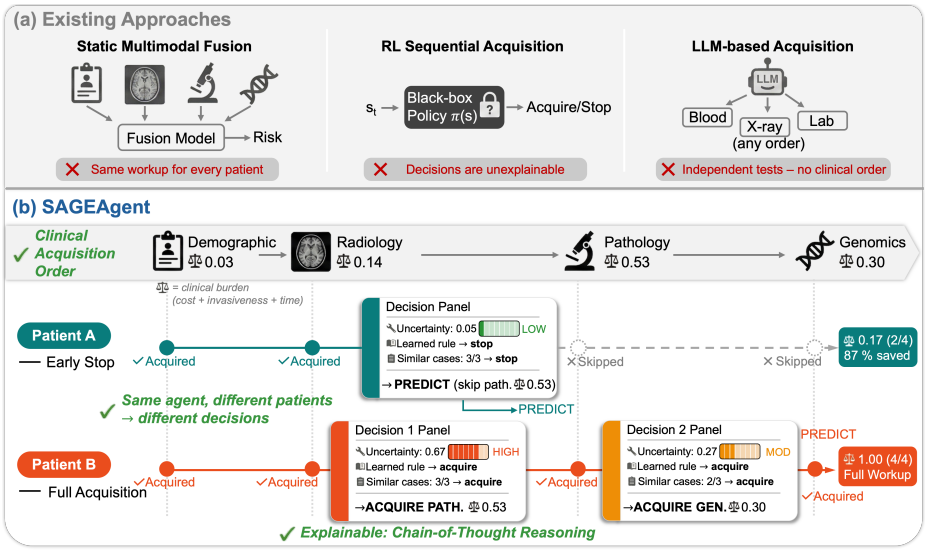


Fig. 1. Overview of modality acquisition strategies. (a) Existing approaches: static fusion applies all modalities uniformly, RL learns adaptive policies without explanation, and LLM-based methods reason over independent tests without clinical ordering. (b) SAGEAgent follows a clinically acquisition order, deciding at each stage whether to acquire or stop based on clinical tools, learned rules, and similar past cases. Patient A stops early (83% burden saved); Patient B proceeds through all modalities.

time, but treat prediction as a static problem—the model outputs a risk score from whatever is observed, without reasoning about **whether the next diagnostic step is worth it for this specific patient**.

Sequential test acquisition has been studied from two directions. Reinforcement learning (RL) methods [2,20] train cost-aware acquisition policies but produce black-box decisions that limit clinical trust. Large language model (LLM) based approaches [13,18,1] offer transparent reasoning but operate on independent diagnostic tests without accounting for the sequential dependencies inherent in clinical modality acquisition. Neither direction models the clinically mandated ordering (see Fig. 1(a) for a comparison).

We propose SAGEAgent (**S**equential **A**cquisition **G**uided by **E**xperience), a training-free, self-evolving LLM-based framework for cost-aware sequential modality acquisition in multimodal survival prediction (Fig. 1(b)). At each stage of the clinical order (demographics → radiology → pathology → genomics), SAGEAgent decides whether to acquire or stop using three signal sources: clinical tools that translate numerical predictions into text, an episodic memory that retrieves similar past cases, and a semantic memory that accumulates decision rules through periodic self-reflection—all without gradient updates. Evaluated on a glioma cohort combining TCGA-LGG [19], TCGA-GBM [21], and BraTS [17] with nested 5×5 cross-validation, SAGEAgent achieves competitive survival pre-

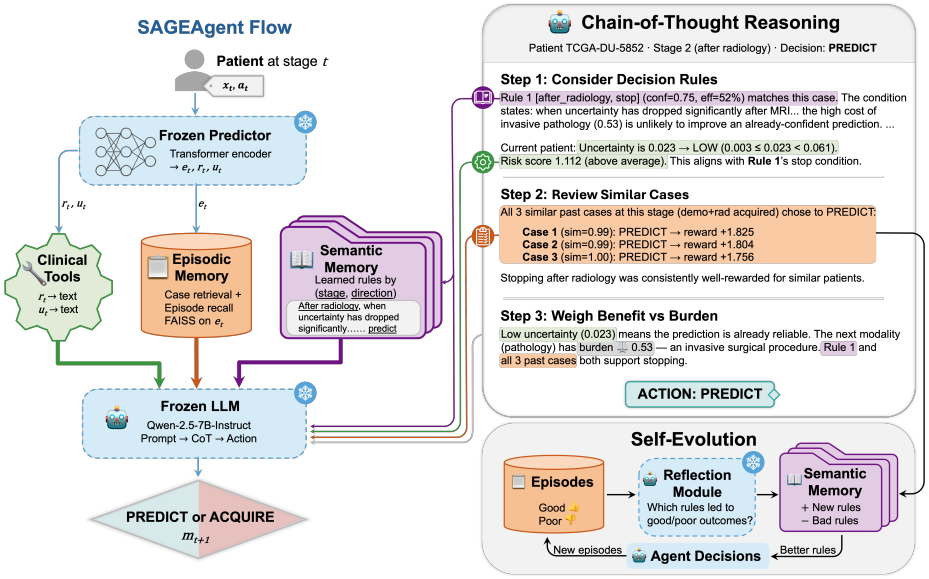


Fig. 2. SAGEAgent architecture and reasoning. SAGEAgent Flow: At each stage, the frozen predictor produces the patient embedding, risk score, and calibrated uncertainty. Three signal sources feed the frozen LLM: clinical tools convert numerical outputs to text, episodic memory retrieves similar patients and past decisions, and semantic memory provides learned decision rules. **Chain-of-Thought Reasoning**: A real reasoning trace after radiology, with highlights tracing each step to its source component. The agent synthesizes all three signals and decides to stop, skipping invasive pathology. **Self-Evolution**: Episodes accumulate with outcome labels; the reflection module analyzes which rules led to good or poor outcomes and updates semantic memory accordingly, forming a closed loop without gradient updates.

diction accuracy while reducing acquisition burden by 55% compared to full workup, providing transparent reasoning for each patient's acquisition pathway.

Our contributions are threefold:

1. **At the learning framework level**, we propose SAGEAgent, a novel framework that formulates cost-aware modality acquisition for multimodal survival prediction as a sequential decision-making problem, while explicitly respecting the clinical ordering of diagnostic procedures.
2. **At the training strategy level**, we introduce a dual-memory self-evolution mechanism that autonomously discovers decision rules from experience, eliminating the need for gradient-based updates or hand-crafted priors.
3. **At the interpretability level**, we provide fully auditable reasoning pathways in SAGEAgent, where each acquisition decision is grounded in explicit chain-of-thought reasoning, thereby bridging cost-aware modality acquisition with clinical transparency.

2 Method

2.1 Problem Formulation

Consider a patient with N diagnostic modalities $\{m_1, \ldots, m_N\}$ arranged in a clinically mandated acquisition order $m_1 \to m_2 \to \cdots \to m_N$. For glioma, this order is demographics → radiology → pathology → genomics, reflecting real clinical dependencies [16]. Each modality m_i carries a clinical burden $b(m_i) \in [0, 1]$, quantified via multi-criteria decision analysis (MCDA) [11] over monetary cost, turnaround time, patient invasiveness, and infrastructure requirements (details in Sec. 3.1). At decision stage $t \in \{1, \ldots, N-1\}$, the agent has already acquired the first t modalities. The decision is binary: acquire m_{t+1} (incurring burden $b(m_{t+1})$), or stop and predict. The goal is to maximize prediction accuracy while minimizing total acquisition burden $B_t = \sum_{i=1}^{t} b(m_i)$. Fig. 2 provides an overview of the proposed framework.

2.2 Multimodal Predictor with Calibrated Uncertainty

Survival predictor. The prediction backbone is a Transformer-based multimodal encoder that handles missing modalities via learnable mask tokens [9,8]. It maps each patient's available modalities to a fused embedding $\mathbf{e}_t \in \mathbb{R}^d$, from which a Cox proportional hazards head [7] produces a scalar risk score r_t. The predictor is trained with modality dropout and **frozen during agent operation**. Architecture and training details are in Sec. 3.2.

Calibrated uncertainty head. On top of the frozen predictor, we train a lightweight uncertainty head that estimates whether the current prediction would change if more modalities were acquired. Given the embedding $\mathbf{e}_t$ and acquisition mask $\mathbf{a}_t$, it outputs a calibrated probability:

$$u_t = \sigma\big(f_\phi([\mathbf{e}_t, \mathbf{a}_t])\big) \tag{1}$$

where f_ϕ is a lightweight MLP and σ is the sigmoid function. The training label is 1 when the absolute difference between the current and full-modality risk scores exceeds a threshold τ, and 0 otherwise. Training data is generated by enumerating clinical-order prefix masks on complete-modality patients. The resulting $u_t \in [0, 1]$ is patient-specific and mask-specific, giving the agent a calibrated signal of whether additional data is likely to matter.

2.3 SAGEAgent: LLM-Based Acquisition Agent

SAGEAgent is built on a frozen LLM that receives no gradient updates. At each stage, it gathers diagnostic signals, retrieves relevant experience, and outputs a binary decision (acquire or predict) with a chain-of-thought explanation. Its behavior improves solely through two memory systems following the episodic–semantic distinction in cognitive science [24].

Clinical decision tools. Two tools translate the frozen predictor's numerical outputs into text. The *Uncertainty Tool* reports u_t from Eq. 1 with a categorical

level (very low to very high) based on quintile thresholds from the training distribution. The *Survival Predictor Tool* reports the risk score r_t and its position relative to the training cohort (below/around/above average). Both tools present outputs as natural-language sentences so that all reasoning occurs in text space.

Episodic memory. The episodic memory provides case-based reasoning through two complementary retrieval mechanisms. *Case retrieval* draws from a FAISS-indexed [10] database of training patients, returning the k nearest neighbors by cosine similarity on ℓ_2-normalized embeddings together with their event status (survived or deceased) and survival time. This provides a static clinical reference: which training patients resemble the current one, and how long did they survive. *Episode recall* draws from the agent's own past decisions. When the agent acquires m_{t+1}, it receives a stage reward $R_{\text{stage}} = -b(m_{t+1}) + \max(0,\, u_t - u_{t+1})$ that penalizes cost and rewards uncertainty reduction. When the agent stops at stage t, a terminal reward $R_{\text{term}} = (1 - u_t) - B_t$ favors early stopping at low uncertainty. Retrieval is stage-matched and re-ranked by the sum of cosine similarity $\text{sim}(\mathbf{e}_t, \mathbf{e}_j)$ and min-max normalized episode reward $\tilde{r}_j \in [0, 1]$, returning the top-k episodes, including the action taken and its outcome.

Semantic memory and self-evolution. The semantic memory stores interpretable decision rules discovered through periodic self-reflection. Each rule is indexed by acquisition stage t and action direction $d \in \{\text{stop}, \text{acquire}\}$. After each episode, we record whether the agent followed or violated each relevant rule. Every N_r episodes (one reflection cycle), episodes are ranked by total reward, and those in the top and bottom quartiles are labeled good and poor respectively. Rule effectiveness is the fraction of followed episodes that led to good outcomes. The reflection module receives the effectiveness report and proposes new rules, which are verified for direction clarity and novelty. Rules that consistently lead to poor outcomes are deprecated, forming a closed loop where rules are discovered, validated, and either reinforced or removed.

3 Experiments & Results

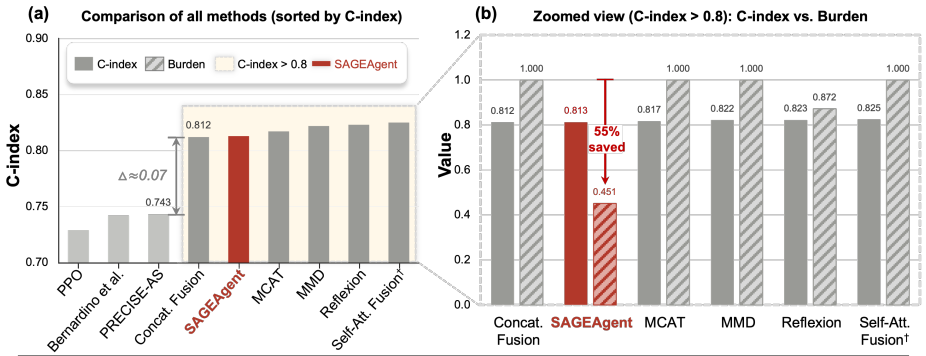
3.1 Dataset and Evaluation Protocol

Dataset and split. We evaluate on a glioma cohort of 962 patients (803 TCGA-LGG/GBM [19,21] and 159 BraTS [17]) spanning four diagnostic modalities in clinical order: demographics, radiology, pathology, and genomics, each encoded into a 32-dimensional feature vector by pretrained extractors [8]. Of the 962 patients, 170 possess all four modalities and form the evaluation set. We adopt nested 5×5 cross-validation. The outer loop splits 170 patients into 5 folds (136/34) for agent training and evaluation. Each inner 5-fold trains the predictor on the 136 patients plus 792 partial-modality patients as augmentation via modality dropout, yielding 25 independent pipelines. Per-patient decisions are aggregated by majority vote ($\geq$3/5 inner models). We report mean ± std across 5 outer folds.

Clinical burden. We quantify modality cost via MCDA [11] over monetary cost, turnaround time, invasiveness, and infrastructure requirements. After nor-

malization, per-modality burdens are 0.03 (demographics), 0.14 (radiology), 0.53 (pathology), and 0.30 (genomics), totaling B=1.0 for a full workup.
Metrics. We report C-index [14], average burden, and hypervolume HV = $(c - 0.5) \times (1.0 - b)$ which captures the accuracy–cost trade-off as a single scalar.

Table 1. Core results on the glioma survival cohort (170 complete-modality patients). Top (a) C-index in ascending order. RL-based methods form a lower cluster around 0.73, while the remaining six methods achieve C-index above 0.8. Top (b) C-index (solid) and burden (hatched) for the six methods with C-index > 0.8, where SAGEAgent saves the most burden. Bottom: per-method C-index and burden as the mean over 5 outer folds with 95% bootstrap confidence intervals (1000 stratified resamples). Bold marks the lowest burden among methods with C-index > 0.8. D, R, P, G denote demographics, radiology, pathology, and genomics, and ■ = 170 patients, ■ = 5 patients, with shades varying linearly between these values. † = predictor backbone used by SAGEAgent.



Category	Method	C-index↑	Burden↓	D	R	P	G
Static Fusion	Concat. Fusion	0.812 [.77, .86]	1.000 [1.00, 1.00]	■	■	■	■
	MCAT [6]	0.817 [.78, .86]	1.000 [1.00, 1.00]	■	■	■	■
	MMD [8]	0.822 [.78, .86]	1.000 [1.00, 1.00]	■	■	■	■
	Self-Att. Fusion†	0.825 [.78, .87]	1.000 [1.00, 1.00]	■	■	■	■
RL-based Acquisition	Bernardino et al. [2]	0.742 [.69, .79]	0.125 [.10, .15]	■	■	■	■
	PRECISE-AS [20]	0.743 [.69, .79]	0.127 [.10, .16]	■	■	■	■
	PPO [22]	0.729 [.68, .78]	0.124 [.10, .16]	■	■	■	■
LLM-based Acquisition	Reflexion [23]	0.823 [.78, .87]	0.872 [.84, .90]	■	■	■	■
	SAGEAgent	0.813 [.77, .86]	**0.451** [.41, .50]	■	■	■	■

3.2 Implementation Details

Predictor. The backbone is a 2-layer Transformer encoder (d=128, 4 heads) with per-modality MLP projections from $\mathbb{R}^{32}$ to $\mathbb{R}^{128}$ and learnable mask tokens for missing modalities, trained with Cox partial likelihood, reconstruction,

Table 2. Component ablation. Each row adds one component to the Base LLM. All configurations use frozen Qwen-2.5-7B-Instruct with majority-vote aggregation. C-index, burden, and HV are reported as the mean over 5 outer folds with 95% bootstrap confidence intervals, following the protocol of Table 1. Bold indicates best burden and HV. D, R, P, G denote demographics, radiology, pathology, and genomics, and ■ = 170 patients, ■ = 5 patients, with shades varying linearly.

Configuration	C-index↑	Burden↓	HV↑	D	R	P	G
Base LLM	0.827 [.79, .87]	0.798 [.77, .82]	0.066 [.06, .08]	■	■	■	■
+ Tools	0.830 [.79, .87]	0.843 [.82, .87]	0.052 [.04, .06]	■	■	■	■
+ Tools + Episodic	0.814 [.77, .86]	0.611 [.56, .66]	0.122 [.10, .14]	■	■	■	■
+ Tools + Semantic	0.819 [.78, .86]	0.706 [.66, .75]	0.094 [.08, .11]	■	■	■	■
SAGEAgent	0.813 [.77, .86]	**0.451** [.41, .50]	**0.172** [.15, .20]	■	■	■	■

and alignment losses under 50% modality dropout (∼458K parameters). The calibrated uncertainty head (Eq. 1) adds ∼6.5K parameters trained with BCE loss, with τ=0.3 set by calibration quality (AUROC = 0.908, ECE = 0.110) on inner-fold validation after post-hoc temperature scaling [12].

SAGEAgent uses Qwen-2.5-7B-Instruct [25] with fully frozen weights, running on a single NVIDIA A6000 GPU. During experience accumulation, each training patient is processed for 3 episodes, yielding ∼408 episodes per fold. Episodic memory retrieves k=3 stage-matched neighbors via FAISS [10]. Semantic memory maintains up to 10 active rules with reflection triggered every 10 patients.

3.3 Core Results

Table 1 reports all methods under the nested 5×5 protocol of Sec. 3.1.

RL methods reduce cost at the expense of accuracy. A C-index gap of approximately 0.07 separates the nine methods into two clusters.(Table 1 top (a)) All three RL baselines reduce burden to ∼0.12 but drop C-index to ∼0.73, effectively discarding informative modalities. Static fusion methods require full workup by design, and Reflexion reduces burden by only 13%.

SAGEAgent reduces burden by 55% at no significant accuracy cost. As shown in Table 1 top (b), SAGEAgent reaches a C-index of 0.813 while reducing burden to 0.451. Its C-index gap to every method with C-index above 0.8 is not statistically significant, with paired 95% bootstrap confidence intervals containing 0 and $p \geq 0.30$, whereas its burden reduction over those methods is significant at $p < 0.001$. SAGEAgent thus offers the strongest accuracy-to-cost balance among the methods compared, rather than the highest accuracy alone.

3.4 Ablation Study

Table 2 isolates the contribution of each component by progressively adding tools, episodic memory, and semantic memory to the frozen LLM. Adding tools alone raises burden from 0.798 to 0.843 because without past experience to

Table 3. Rules discovered by semantic memory across five folds. Eff. = fraction of rule-following episodes with top-quartile reward (random baseline = 0.25). n = total episodes where the rule was followed across folds. Stability indicates whether the rule, once discovered, was ever deprecated and re-proposed during training.

Discovered Rule	Folds	n	Eff.	Stability
Stopping rules				
"After radiology, predict now ... risk error is small and pathology cost outweighs potential gain."	5/5	876	.46–.52	Stable
"After pathology, predict now ... prediction quality is high and genomics cost is not justified."	5/5	387	.39–.62	Stable
Acquisition rules				
"After demographics, acquire radiology ... uncertainty is high and imaging cost is low."	4/5	493	.38–.67	Stable
"After radiology, acquire pathology ... uncertainty remains high and risk error is large."	4/5	353	.44–.58	Churned

contextualize uncertainty and risk signals, the LLM defaults to acquiring more. Episodic memory provides the largest individual burden reduction (0.843→0.611) by surfacing cases where early stopping succeeded, while semantic memory offers a complementary effect (0.843→0.706) through interpretable stopping criteria. Combined, SAGEAgent achieves the lowest burden (0.451) and highest HV (0.172), a 2.6× improvement over the Base LLM, whose marginally higher C-index is not significant. A naive policy that stops once u_t falls below a fixed τ reaches C-index 0.774, 0.769, and 0.750 at τ=0.3, 0.5, and 0.7 with burden 0.347, 0.235, and 0.144, which is 0.04 to 0.06 below SAGEAgent. The tools and dual memory therefore add value beyond the uncertainty head.

3.5 Self-Evolution Analysis

The ablation in Sec. 3.4 shows that semantic memory reduces burden, but does not reveal what rules it learns. We now examine whether the discovered rules are consistent across folds. Each outer fold runs 10 reflection cycles from an empty rule set with no shared initialization.

Semantic memory discovers consistent rules across folds. Table 3 lists the four rules emerging in at least four of five folds, all achieving effectiveness well above the 0.25 random baseline and collectively covering over 2,100 decisions. Each fold generates 10.8 candidate rules on average, of which 4.6 survive pruning.

Stopping rules are easier to learn than acquisition rules. The two stopping rules emerge in the first two reflection cycles and are never deprecated. In contrast, the acquisition rule "after radiology, acquire pathology" accumulates 13 deprecated instances across folds before a stable version survives in four of five. Reflection logs reveals a two-phase pattern: cycles 1–2 establish core stopping

rules that persist unchanged, while cycles 3–10 are dominated by acquire-rule exploration where most proposals are deprecated due to near-zero effectiveness.

4 Conclusion

We presented SAGEAgent, a training-free LLM agent for cost-aware sequential modality acquisition in survival prediction. On a glioma cohort it reduces acquisition burden by 55% with no statistically significant loss in C-index, grounding each decision in auditable reasoning. Notably, it learns to skip invasive pathology for half of the cohort, suggesting that many glioma patients may not require biopsy for reliable prognostication. Self-evolution analysis shows that stopping rules converge reliably while acquisition rules are harder to learn. Our evidence is limited to a single retrospective glioma cohort with complete four-modality data, so external multi-site validation, other cancer types, and prospective clinical study remain future work.

Acknowledgments. This research was supported by the National Institutes of Health (NIH) through grants F30CA275020, R01CA253923 (Landman & Maldonado), R01CA275015 (Maldonado & Lenburg), U01CA152662 (Grogan), U01CA196405 (Maldonado), and P30CA068485-29S1, as well as the National Science Foundation (NSF) through CAREER 1452485 and grant 2040462. Additional support was provided by the Vanderbilt Institute for Surgery and Engineering through T32EB021937-07, the Vanderbilt Institute for Clinical and Translational Research via UL1TR002243-06, the Pierre Massion Directorship in Pulmonary Medicine, and the American College of Radiology Fund for Collaborative Research in Imaging (FCRI) Grant.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bani-Harouni, D., Pellegrini, C., Özsoy, E., Keicher, M., Navab, N.: Language agents for hypothesis-driven clinical decision making with reinforcement learning. arXiv preprint arXiv:2506.13474 (2025)
2. Bernardino, G., Jonsson, A., Loncaric, F., Martí Castellote, P.M., Sitges, M., Clarysse, P., Duchateau, N.: Reinforcement learning for active modality selection during diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 592–601. Springer (2022)
3. Boehm, K.M., Khosravi, P., Vanguri, R., Gao, J., Shah, S.P.: Harnessing multimodal data integration to advance precision oncology. Nature Reviews Cancer **22**(2), 114–126 (2022)
4. Braman, N., Gordon, J.W., Goossens, E.T., Willis, C., Stumpe, M.C., Venkataraman, J.: Deep orthogonal fusion: multimodal prognostic biomarker discovery integrating radiology, pathology, genomic, and clinical data. In: International conference on medical image computing and computer-assisted intervention. pp. 667–677. Springer (2021)

5. Chen, R.J., Lu, M.Y., Wang, J., Williamson, D.F., Rodig, S.J., Lindeman, N.I., Mahmood, F.: Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. IEEE Transactions on Medical Imaging **41**(4), 757–770 (2020)
6. Chen, R.J., Lu, M.Y., Weng, W.H., Chen, T.Y., Williamson, D.F., Manz, T., Shady, M., Mahmood, F.: Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4015–4025 (2021)
7. Cox, D.R.: Regression models and life-tables. Journal of the royal statistical society: Series B (methodological) **34**(2), 187–202 (1972)
8. Cui, C., Liu, H., Liu, Q., Deng, R., Asad, Z., Wang, Y., Zhao, S., Yang, H., Landman, B.A., Huo, Y.: Survival prediction of brain cancer with incomplete radiology, pathology, genomic, and demographic data. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 626–635. Springer (2022)
9. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers). pp. 4171–4186 (2019)
10. Douze, M., Guzhva, A., Deng, C., Johnson, J., Szilvasy, G., Mazaré, P.E., Lomeli, M., Hosseini, L., Jégou, H.: The faiss library. IEEE Transactions on Big Data (2025)
11. Guitouni, A., Martel, J.M.: Tentative guidelines to help choosing an appropriate mcda method. European journal of operational research **109**(2), 501–521 (1998)
12. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On calibration of modern neural networks. In: International conference on machine learning. pp. 1321–1330. PMLR (2017)
13. Hager, P., Jungmann, F., Holland, R., Bhagat, K., Hubrecht, I., Knauer, M., Vielhauer, J., Makowski, M., Braren, R., Kaissis, G., et al.: Evaluation and mitigation of the limitations of large language models in clinical decision-making. Nature medicine **30**(9), 2613–2622 (2024)
14. Harrell, F.E., Califf, R.M., Pryor, D.B., Lee, K.L., Rosati, R.A.: Evaluating the yield of medical tests. Jama **247**(18), 2543–2546 (1982)
15. Lipkova, J., Chen, R.J., Chen, B., Lu, M.Y., Barbieri, M., Shao, D., Vaidya, A.J., Chen, C., Zhuang, L., Williamson, D.F., et al.: Artificial intelligence for multimodal data integration in oncology. Cancer cell **40**(10), 1095–1110 (2022)
16. Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., Hawkins, C., Ng, H., Pfister, S.M., Reifenberger, G., et al.: The 2021 who classification of tumors of the central nervous system: a summary. Neuro-oncology **23**(8), 1231–1251 (2021)
17. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014)
18. Nori, H., Daswani, M., Kelly, C., Lundberg, S., Ribeiro, M.T., Wilson, M., Liu, X., Sounderajah, V., Carlson, J., Lungren, M.P., et al.: Sequential diagnosis with language models. arXiv preprint arXiv:2506.22405 (2025)
19. Pedano, N., Flanders, A.E., Scarpace, L., Mikkelsen, T., Eschbacher, J.M., Hermes, B., Ostrom, Q.: Radiology data from the cancer genome atlas low grade glioma [tcga-lgg] collection. The cancer imaging archive **2** (2016)

20. Saadat, A., Hashemi, N., Vaseli, H., Tsang, M.Y., Luong, C., Van de Panne, M., Tsang, T.S., Abolmaesumi, P.: Precise-as: Personalized reinforcement learning for efficient point-of-care echocardiography in aortic stenosis diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 474–484. Springer (2025)
21. Scarpace, L., Mikkelsen, T., Cha, S., Rao, S., Tekchandani, S., Gutman, D., Saltz, J.H., Erickson, B.J., Pedano, N., Flanders, A.E., et al.: The cancer genome atlas glioblastoma multiforme collection (tcga-gbm). The Cancer Imaging Archive (2016)
22. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347 (2017)
23. Shinn, N., Cassano, F., Gopinath, A., Narasimhan, K., Yao, S.: Reflexion: Language agents with verbal reinforcement learning. Advances in neural information processing systems **36**, 8634–8652 (2023)
24. Tulving, E., et al.: Episodic and semantic memory. Organization of memory **1**(381-403), 1 (1972)
25. Yang, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Li, C., Liu, D., Huang, F., Wei, H., et al.: Qwen2. 5 technical report. arXiv e-prints pp. arXiv–2412 (2024)