Retrieval-based Mixture-of-Experts for Patient-Specific Cancer Survival Prediction with Incomplete Multimodal Data

Minjoo Lim[1], Bogyeong Kang[1], Hyun Jung Lee[1], Eunjung Jo[1], Keun-Soo Heo[1], Mingxia Liu[2⋆], and Tae-Eui Kam[1⋆]

[1] Department of Artificial Intelligence, Korea University, Republic of Korea.
{imj700, kamte}@korea.ac.kr
[2] Department of Radiology and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
mingxia_liu@med.unc.edu

Abstract. Survival prediction is crucial in clinical oncology for cancer prognosis and treatment planning. Recent survival prediction models have shown that integrating pathology whole-slide images (WSIs) and gene expression data effectively improves performance by leveraging complementary morphological and molecular information. However, in clinical practice, complete multimodal data are often unavailable due to high costs, privacy concerns, and workflow constraints, resulting in substantial modality missingness. In such settings, existing methods typically **(i)** rely on modality-specific or modality-common information shared across patients, often overlooking patient-specific cross-modal dependencies, and **(ii)** adopt a single static prediction pathway, which limits their ability to adapt to diverse missingness patterns. To address these challenges, we propose Retrieval-based Mixture-of-Experts (**RMoE**) for robust cancer survival prediction with incomplete multimodal data. To explicitly capture patient-specific cross-modal dependencies, RMoE utilizes *bidirectional cross-modal retrieval*, identifying phenotypically similar patients and enabling patient-adaptive imputation through biologically relevant neighbors. Furthermore, we introduce a *reliability-aware Mixture-of-Experts* architecture, where a router dynamically weights specialized experts by assessing retrieval confidence and modality availability, ensuring stable predictions regardless of diverse missing modality settings. Experiments on three public cancer datasets with 60% missingness show that RMoE consistently outperforms current state-of-the-art methods.

Keywords: Cancer Survival Prediction · Missing Modality · Retrieval · Mixture-of-Experts

1 Introduction

Survival prediction is a fundamental task in clinical oncology for cancer prognosis and personalized treatment planning [1, 16]. Pathology whole-slide images

⋆Co-corresponding authors.

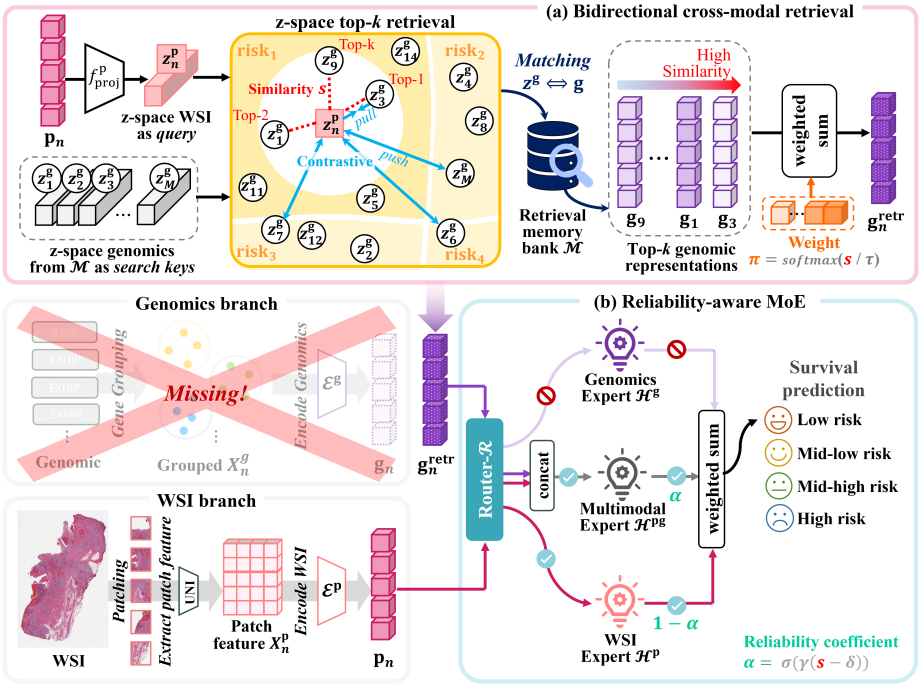

Fig. 1. Overview of RMoE. While RMoE can handle both genomics-missing $\{\mathbf{p}_n, \varnothing\}$ and WSI-missing $\{\varnothing, \mathbf{g}_n\}$ cases, this figure depicts the genomics-missing case for illustration purposes. **(a)** Bidirectional cross-modal retrieval imputes the missing representation ($\mathbf{g}_n^{\text{retr}}$ or $\mathbf{p}_n^{\text{retr}}$) by using the available representation ($\mathbf{p}_n$ or $\mathbf{g}_n$) as a query to retrieve clinically similar patients from a retrieval memory bank $\mathcal{M}$ based on cosine similarity s in a shared z-space. **(b)** Reliability-aware MoE adaptively interpolates between unimodal ($\mathcal{H}^{\text{p}}, \mathcal{H}^{\text{g}}$) and multimodal experts ($\mathcal{H}^{\text{pg}}$) based on the reliability coefficient α, allowing the model to trust the retrieved modality at high confidence.

(WSIs) and gene expression data are two key modalities in this field, providing complementary morphological and molecular cues that characterize the tumor microenvironment [6,12,31]. Recent deep learning studies have explored survival prediction by integrating these data through sophisticated cross-modal fusion strategies [4–6,25]. Despite their strong performance, most existing multimodal models are built on the ideal assumption of complete data availability for all patients. However, in clinical practice, acquiring both modalities is often precluded by high data acquisition costs, privacy concerns, and fragmented clinical workflows [17–19,23]. Consequently, multimodal datasets are often incomplete, limiting the applicability of models that assume full data availability.

To address missing modalities, prior works have explored generative imputation [2,11,27], shared representation learning [14,21,30], and latent distillation techniques [22]. In survival analysis, these approaches have been extended by

replacing missing modalities with learned prototypes [13] or by Large Language Models (LLMs) to model modality-specific and shared distributions [26]. Most of these methods mainly rely on modality-specific or modality-common information shared across patients. As such, they often fail to capture patient-specific cross-modal dependencies that are critical for individualized survival prediction. Moreover, these methods typically use a static prediction pathway shared across all missingness scenarios, which struggles to adapt to diverse and complex patterns of modality availability encountered in clinical workflows.

To this end, we propose Retrieval-based Mixture-of-Experts (**RMoE**) for robust survival prediction with incomplete multimodal data, as shown in Fig. 1. RMoE utilizes *bidirectional cross-modal retrieval* (WSI-to-genomics and genomics-to-WSI retrieval) to perform patient-adaptive imputation, explicitly capturing patient-specific cross-modal dependencies even when modalities are missing. Specifically, the available modality is used to identify phenotypically similar patients and the missing modality is then imputed via similarity-weighted aggregation of the corresponding representations from nearest neighbors. To ensure biologically meaningful retrieval, the shared embedding space is optimized by aligning cross-modal embeddings of patients with similar survival risks. RMoE further incorporates a *reliability-aware Mixture-of-Experts (MoE)* architecture that dynamically allocates experts specialized for different modality configurations, rather than sharing a single prediction pathway across all missingness patterns. A router module adaptively weights unimodal and multimodal experts based on retrieval confidence, enabling robust prediction across different modality availability patterns. Experiments on three datasets with 60% modality missingness demonstrate that RMoE outperforms current state-of-the-art methods.

2 Methodology

Problem Formulation. Given a dataset of N patients, we denote the n-th patient as $\mathbf{X}_n = (X_n^{\mathrm{p}}, X_n^{\mathrm{g}})$, where X_n^{p} is a bag of WSI patch features and X_n^{g} is a bag of genomic pathway values. We encode each modality into a fixed-dimensional representation using modality-specific encoders: $\mathbf{p}_n = \mathcal{E}^{\mathrm{p}}(X_n^{\mathrm{p}})$ and $\mathbf{g}_n = \mathcal{E}^{\mathrm{g}}(X_n^{\mathrm{g}})$. Survival labels are defined as (c_n, t_n), where $c_n \in \{0, 1\}$ is the censoring status ($c_n = 1$ if right-censored) and $t_n \in \mathbb{R}^+$ denotes the overall survival time in months. Following standard practice in survival analysis [6,8,28], the survival time is discretized into I intervals, and the model predicts a set of hazard probabilities $(h_n^{(1)}, \ldots, h_n^{(I)})$. For each patient n, the representation set is denoted as $\{\mathbf{p}_n, \mathbf{g}_n\}$ with both modalities available, whereas it simplifies to $\{\mathbf{p}_n, \varnothing\}$ or $\{\varnothing, \mathbf{g}_n\}$ when genomics or WSI data are absent, respectively.

Method Overview. RMoE addresses modality missingness through two key components: a bidirectional cross-modal retrieval and a reliability-aware MoE, as illustrated in Fig. 1. Given a sample with incomplete modalities, RMoE performs bidirectional cross-modal retrieval using the available modality to retrieve phenotypically similar neighbors from a memory bank, which are used to impute the missing modality representation. The retrieved representation is then inte-

grated through a reliability-aware MoE, where a routing function computes a retrieval reliability coefficient to dynamically weight unimodal and multimodal experts. By adaptively controlling the influence of retrieved information, RMoE ensures effective survival prediction under diverse missingness patterns.

1) Bidirectional Cross-Modal Retrieval. To impute missing modalities, we maintain a *retrieval memory bank* $\mathcal{M}$ with multimodal patient entries from the training set. Each entry in $\mathcal{M}$ for patient j consists of: **(i)** *modality-specific representations* $\mathbf{p}_j, \mathbf{g}_j \in \mathbb{R}^D$ for survival prediction, **(ii)** *search keys* $\mathbf{z}_j^{\mathrm{p}}, \mathbf{z}_j^{\mathrm{g}} \in \mathbb{R}^{d_z}$ which are projected into a shared embedding space for retrieval, **(iii)** a risk-group label $\mathbf{risk}_j$, determined by the index of the discretized survival-time interval, and **(iv)** patient identifier $\mathbf{idx}_j$. The *search keys* are obtained by modality-specific projection heads, $\mathbf{z}_j^{\mathrm{p}} = f_{\mathrm{proj}}^{\mathrm{p}}(\mathbf{p}_j)$ and $\mathbf{z}_j^{\mathrm{g}} = f_{\mathrm{proj}}^{\mathrm{g}}(\mathbf{g}_j)$, and are ℓ_2-normalized for cosine similarity retrieval. Mathematically, each memory entry is defined as $\mathcal{M}_j = (\mathbf{p}_j, \mathbf{g}_j, \mathbf{z}_j^{\mathrm{p}}, \mathbf{z}_j^{\mathrm{g}}, \mathbf{risk}_j, \mathbf{idx}_j)$, and $\mathcal{M}$ is updated every epoch. Note that we distinguish *search keys* $\mathbf{z}$ from $\mathbf{p}$ and $\mathbf{g}$, where $\mathbf{z}$ is learned for similarity-based cross-modal retrieval in a shared embedding space, while $\mathbf{p}$ and $\mathbf{g}$ are preserved for survival prediction. This decoupled approach allows the shared embedding space to be optimized for similarity-based alignment without distorting the original modality-specific representations required for survival prediction.

Given an input with a missing modality, we retrieve the missing modality representation from $\mathcal{M}$ in a bidirectional manner, *i.e.* both WSI-to-genomics and genomics-to-WSI. As shown in Fig. 1(a), when the observed input is $\{\mathbf{p}_n, \varnothing\}$, we compute a *query* $\mathbf{z}_n^{\mathrm{p}}$ for WSI-to-genomics retrieval. To retrieve clinically relevant neighbors, we restrict memory candidates to entries with the same risk-group label $\mathbf{risk}_j$ during training. This restriction acts as a clinical prior, ensuring that the retrieved representations are not just mathematically similar but also clinically relevant to the patient's prognosis. We then retrieve the missing genomic representation $\mathbf{g}_n^{\mathrm{retr}}$ from $\mathcal{M}$ as a weighted sum of top-k neighbors:

$$\mathbf{g}_n^{\mathrm{retr}} = \sum\nolimits_{j \in \mathcal{N}_k} \pi_j \mathbf{g}_j, \quad \pi_j = \mathrm{softmax}\Big(\frac{s_j}{\tau}\Big), \quad s_j = (\mathbf{z}_n^{\mathrm{p}})^\top \mathbf{z}_j^{\mathrm{g}}, \tag{1}$$

where τ is the temperature parameter, $\mathcal{N}_k$ denotes the top-k entries ranked by similarity s_j in the shared z-space, and their corresponding $\mathbf{g}_j$ are aggregated. By aggregating multiple nearest neighbors, complementary information across retrieved samples mitigates the potential noise or variability inherent in any individual instance, leading to more stable modality completion. When the observed input is $\{\varnothing, \mathbf{g}_n\}$, we perform genomics-to-WSI retrieval. Following a symmetric procedure, we compute the *query* $\mathbf{z}_n^g$ and leverage the same retrieval memory bank $\mathcal{M}$ to retrieve the missing WSI representation $\mathbf{p}_n^{\mathrm{retr}}$. After retrieval, we can complete the representation set by imputing the missing modality with the retrieved representation, *e.g.*, $\{\mathbf{p}_n, \varnothing\} \rightarrow \{\mathbf{p}_n, \mathbf{g}_n^{\mathrm{retr}}\}$ and $\{\varnothing, \mathbf{g}_n\} \rightarrow \{\mathbf{p}_n^{\mathrm{retr}}, \mathbf{g}_n\}$.

Since our retrieval process relies on cosine similarity in the projected $\mathbf{z}$-space, it is crucial to learn a risk-aware embedding space where clinically similar patients are mapped close to each other. To this end, we introduce a risk-aware cross-modal contrastive loss $\mathcal{L}_c$ to optimize the projection heads $f_{\mathrm{proj}}^{\mathrm{p}}$ and $f_{\mathrm{proj}}^{\mathrm{g}}$. Given a WSI *query* $\mathbf{z}_n^{\mathrm{p}}$ and all genomic embeddings $\{\mathbf{z}_j^{\mathrm{g}}\}_{j=1}^M$ stored in $\mathcal{M}$ as *keys*,

we define the positive set as memory entries belonging to the same risk-group: $\mathcal{P}(n) = \{j \mid \mathbf{risk}_j = \mathbf{risk}_n,\ \mathbf{idx}_j \neq \mathbf{idx}_n\}$. To avoid a degenerate shortcut where the query matches its own stored key in the bank, we exclude entries with the same patient identifier from the positive set by including the constraint $\mathbf{idx}_j \neq \mathbf{idx}_n$. Among these positives, we select the top-k entries with the highest similarity, denoted as $\mathcal{P}_k(n)$. Formally, the contrastive loss is defined as

$$\mathcal{L}_c = -\log\left[\sum\nolimits_{j\in\mathcal{P}_k(n)} \exp\Big((\mathbf{z}_n^{\mathrm{p}})^\top \mathbf{z}_j^{\mathrm{g}}/\tau\Big) \Big/ \sum\nolimits_{j=1}^{M} \exp\Big((\mathbf{z}_n^{\mathrm{p}})^\top \mathbf{z}_j^{\mathrm{g}}/\tau\Big)\right]. \tag{2}$$

Conversely, when $\mathbf{z}_n^{\mathrm{g}}$ serves as the *query*, $\{\mathbf{z}_j^{\mathrm{p}}\}_{j=1}^{M}$ are used as *keys* in the contrastive loss term. This contrastive loss encourages cross-modal alignment across different patients within the same risk-group in the shared $\mathbf{z}$-space, thereby improving the reliability of nearest-neighbor retrieval for modality completion.

2) Reliability-Aware Mixture-of-Experts. We adopt three expert heads corresponding to different modality availability settings: a WSI-only expert $\mathcal{H}^{\mathrm{p}}(\cdot)$, a genomics-only expert $\mathcal{H}^{\mathrm{g}}(\cdot)$, and a multimodal expert $\mathcal{H}^{\mathrm{pg}}([\cdot\|\cdot])$, as shown in Fig. 1(b). Each expert output is mapped to logits for survival prediction with a classifier head. When a modality is retrieved from memory, we quantify a retrieval confidence score $s = \max_{j\in\mathcal{N}_k} s_j$, representing the maximum cosine similarity among the top-k retrieved neighbors. The reliability-aware router $\mathcal{R}$ converts s into a mixing weight $\alpha \in (0,1)$: $\alpha = \sigma\big(\gamma(s-\delta)\big)$, where δ is a confidence threshold and γ controls the slope. If one modality is missing, $\mathcal{R}$ interpolates between the unimodal expert and the multimodal expert according to the reliability coefficient α; otherwise, when both modalities are available, $\mathcal{R}$ directly uses the multimodal expert:

$$\hat{\mathbf{y}}_n = \begin{cases} (1-\alpha)\mathcal{H}^{\mathrm{p}}(\mathbf{p}_n) + \alpha\mathcal{H}^{\mathrm{pg}}([\mathbf{p}_n\|\mathbf{g}_n^{\mathrm{retr}}]), & \text{if } \{\mathbf{p}_n, \varnothing\}, \\ (1-\alpha)\mathcal{H}^{\mathrm{g}}(\mathbf{g}_n) + \alpha\mathcal{H}^{\mathrm{pg}}([\mathbf{p}_n^{\mathrm{retr}}\|\mathbf{g}_n]), & \text{if } \{\varnothing, \mathbf{g}_n\}, \\ \mathcal{H}^{\mathrm{pg}}([\mathbf{p}_n\|\mathbf{g}_n]), & \text{if } \{\mathbf{p}_n, \mathbf{g}_n\}. \end{cases} \tag{3}$$

Intuitively, when retrieval confidence is low ($s \ll \delta$), $\alpha \approx 0$ and the prediction is dominated by the unimodal expert. When retrieval confidence is high ($s \gg \delta$), $\alpha \approx 1$ and the model trusts the multimodal expert that integrates the retrieved representation. Note that the final output $\hat{\mathbf{y}}_n$ is an I-dimensional logit vector $(\hat{y}_n^{(1)}, \ldots, \hat{y}_n^{(I)})$ and can be converted into hazard probabilities by $\hat{h}_n^{(i)} = \sigma(\hat{y}_n^{(i)})$.

Objective Function. Given the predicted hazard probabilities $(\hat{h}_n^{(1)}, \ldots, \hat{h}_n^{(I)})$, we define the survival function as $\hat{S}_n^{(i)} = \prod_{u=1}^{i}(1-\hat{h}_n^{(u)})$. The survival prediction loss is then computed for patient n with ground-truth time interval i, following [29]: $\mathcal{L}_{\mathrm{s}} = -\Big[c_n \log \hat{S}_n^{(i)} + (1-c_n)\log \hat{S}_n^{(i-1)} + (1-c_n)\log \hat{h}_n^{(i)}\Big]$. In addition, we introduce an auxiliary alignment loss $\mathcal{L}_a$ that enforces direct pairwise consistency in the projected space. Unlike the contrastive loss $\mathcal{L}_c$, which encourages discriminative alignment by contrasting positive pairs against negatives, $\mathcal{L}_a$ serves as a local constraint to minimize cross-modal distance. Specifically, when both modalities are available, we minimize the cosine distance between their projected z-space embeddings, *i.e.*, $\mathbb{E}\big[1-\cos(\mathbf{z}^{\mathrm{p}}, \mathbf{z}^{\mathrm{g}})\big]$, and when a modality is missing, we align the z-space embedding of the available modality with

its retrieved counterpart, *i.e.*, $\mathbb{E}\big[1 - \cos(\mathbf{z}, \mathbf{z}^{\text{retr}})\big]$. Finally, the overall training objective is: $\mathcal{L}_{total} = \mathcal{L}_s + \lambda_1 \mathcal{L}_c + \lambda_2 \mathcal{L}_a$, where λ_1 and λ_2 balance the contrastive and auxiliary alignment losses relative to the primary survival prediction loss.

3 Experiments

Experimental Setup. We conduct experiments on three public cohorts from The Cancer Genome Atlas (TCGA) [24], with paired pathology WSI and genomics data: Bladder Urothelial Carcinoma (BLCA) with 378 cases, Breast Invasive Carcinoma (BRCA) with 1007 cases, and Uterine Corpus Endometrial Carcinoma (UCEC) with 478 cases. WSI patch extraction is performed using a CLAM [15]-based pipeline, followed by patch encoding with UNI [3]. Genes are grouped into biological pathways following [6, 26]. For the modality-specific encoders, $\mathcal{E}^{\text{p}}$ is implemented using TransMIL [20], and $\mathcal{E}^{\text{g}}$ is implemented using an SNN [7]. We use the Adam optimizer with a learning rate of 2×10^{-4} and a weight decay of 1×10^{-5}. The loss weighting parameters λ_1 and λ_2 are both set to 0.1. For retrieval, we set k to 5 and τ to 0.1. For the reliability coefficient, we use a confidence threshold δ of 0.4 and a slope γ of 10. All models are trained for 50 epochs with a batch size of 1.

We compare RMoE with unimodal and multimodal survival prediction methods. TransMIL [20] and SNN [7] are used as WSI-only and genomics-only methods, respectively. For multimodal approaches trained with complete modalities, we compare against SurvPath [6]. We also compare against state-of-the-art multimodal frameworks that handle missing modalities during both training and evaluation: ShaSpec [21], M3Care [30], MAP [10], ProSurv [13], and DisPro [26]. All methods are evaluated using 5-fold cross-validation, with the data split into training, validation, and test sets in a 6:2:2 ratio. In the training set, one modality is randomly removed with a probability of 60%, where WSI and genomics are removed at equal rates. For validation and testing, performance is evaluated across all modality configurations (WSI-only, genomics-only, and WSI&genomics), using the concordance index (C-Index) with standard deviation (std).

Experimental Results. Table 1 reports the C-Index comparison across three datasets under all modality configurations, demonstrating that our RMoE consistently outperforms competing methods [6, 7, 10, 13, 20, 21, 26, 30], thereby highlighting the effectiveness of our retrieval-based imputation and MoE-driven robustness to different modality availability patterns. Furthermore, RMoE surpasses unimodal methods [7, 20] specifically optimized for single-modality inputs, suggesting that cross-modal learning yields more robust representations. Remarkably, our model also outperforms the multimodal survival prediction model [6] trained with complete-modality samples. In several cases, RMoE even demonstrates higher performance when evaluated with a missing modality than competing methods evaluated with both modalities available. For instance, on the BRCA dataset, our model achieves a C-Index of 0.5902 using only genomics

https://www.cancer.gov/tcga

Table 1. C-Index comparison of our proposed RMoE (Ours) with competing methods on the BLCA, BRCA, and UCEC datasets. "W" and "G" denote WSI and genomics modalities, respectively. • and ○ denote available and missing modalities, respectively. † denotes methods trained using unimodal-only or complete modalities, while the remaining methods explicitly handle missing modalities. Bold indicates the best results, while underline indicates the second-best results.

Modalities		Methods	BLCA	BRCA	UCEC	Average
W	G					
•	○	TransMIL† [20]	0.5004±0.084	0.5421±0.048	0.6152±0.096	0.5526
•	○	ShaSpec [21]	0.5080±0.059	0.5678±0.057	0.5822±0.076	0.5527
•	○	M3Care [30]	0.5275±0.054	0.5649±0.059	0.6464±0.091	0.5796
•	○	MAP [10]	0.4958±0.035	0.5484±0.087	0.5513±0.046	0.5318
•	○	Prosurv [13]	0.5249±0.106	0.5447±0.066	0.6636±0.089	0.5777
•	○	DisPro [26]	0.5390±0.081	0.5674±0.058	0.5924±0.080	0.5663
•	○	RMoE (Ours)	**0.5426±0.027**	**0.5803±0.046**	**0.6637±0.079**	**0.5955**
○	•	SNN† [7]	0.5024±0.038	0.5275±0.055	0.6095±0.068	0.5465
○	•	ShaSpec [21]	0.5980±0.026	0.5374±0.042	0.5877±0.075	0.5744
○	•	M3Care [30]	0.5444±0.060	0.5536±0.052	0.6436±0.091	0.5805
○	•	MAP [10]	0.5890±0.028	0.5844±0.050	0.6365±0.091	0.6033
○	•	Prosurv [13]	0.5729±0.059	0.5414±0.091	0.6260±0.110	0.5881
○	•	DisPro [26]	0.5932±0.047	0.5348±0.074	0.6371±0.083	0.5884
○	•	RMoE (Ours)	**0.6159±0.024**	**0.5902±0.058**	**0.6513±0.073**	**0.6191**
•	•	SurvPath† [6]	0.5770±0.049	0.5858±0.044	0.5971±0.059	0.5866
•	•	ShaSpec [21]	0.5654±0.066	0.5646±0.048	0.5602±0.101	0.5634
•	•	M3Care [30]	0.5700±0.047	0.5563±0.066	0.6715±0.076	0.5993
•	•	MAP [10]	0.5532±0.023	0.5081±0.063	0.5994±0.080	0.5536
•	•	Prosurv [13]	0.5905±0.025	0.5230±0.082	0.6776±0.088	0.6006
•	•	DisPro [26]	0.5783±0.037	0.5842±0.066	0.6151±0.071	0.5925
•	•	RMoE (Ours)	**0.6064±0.022**	**0.6270±0.055**	**0.6911±0.116**	**0.6415**

data, exceeding the multimodal performance of every competing method [6, 10, 13, 21, 26, 30]. By effectively utilizing both complete and incomplete data during training, RMoE is exposed to broader and more diverse patient samples, resulting in improved generalization. Ultimately, our approach achieves strong performance regardless of modality availability. The consistent improvements across datasets and modality configurations validate the effectiveness of our design in addressing real clinical scenarios with incomplete multimodal data.

Effectiveness of Bidirectional Cross-Modal Retrieval. As shown in Fig. 2(a), incorporating the bidirectional cross-modal retrieval improves performance, indicating that our retrieval mechanism effectively compensates for missing modality information by leveraging patient-specific cross-modal dependencies. Fig. 2(d) further visualizes the retrieval process using t-SNE. In both WSI-to-genomics and genomics-to-WSI retrieval, samples from the same risk-group cluster around each query, indicating that our model retrieves prognostically similar cross-modal embeddings. This confirms that the proposed retrieval mechanism, optimized by the contrastive loss, effectively learns cross-modal alignment with missing data.

Effectiveness of Reliability-Aware MoE. In Fig. 2(a), removing the reliability-aware MoE leads to a noticeable degradation, showing that specialized unimodal and multimodal experts enhance robustness under severe missing modality settings. Importantly, our router further leverages retrieval confidence to adaptively weight retrieved representations, enabling more reliable expert allocation. As a

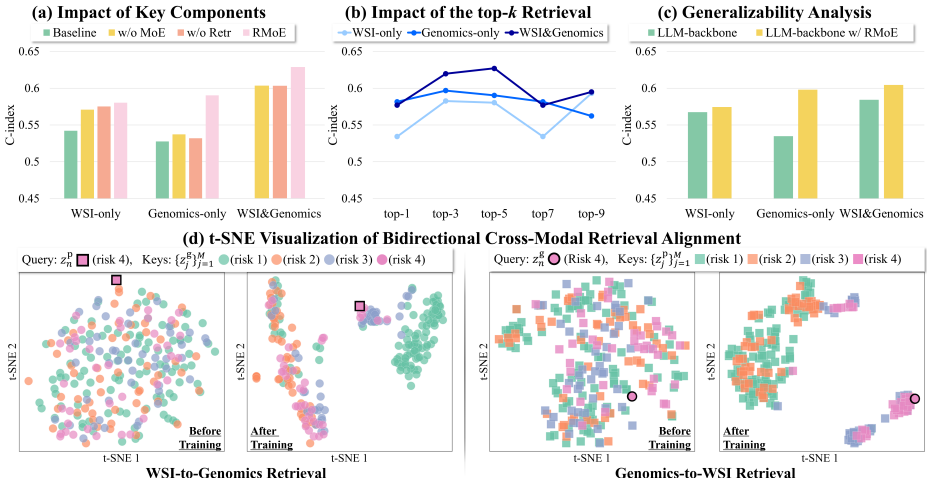

Fig. 2. **(a)** Impact of key components in RMoE. Baseline indicates a WSI-only or genomics-only branch, w/o MoE removes the reliability-aware MoE, while w/o Retr removes the bidirectional cross-modal retrieval. **(b)** Impact of the number of retrieved neighbors (top-1 to top-9). **(c)** Comparison between the LLM-based backbone and its integration with RMoE. **(d)** t-SNE visualization of bidirectional cross-modal retrieval. All results are based on the BRCA dataset.

result, when combined with retrieval (i.e., RMoE), this design achieves the best performance, highlighting complementary effects between retrieval and MoE.

Impact of Top-k Retrieval. We analyze the influence of the number of retrieved top-k neighbors in Fig. 2(b). Using few samples (*e.g.*, top-1) makes the model sensitive to noisy retrievals, while too many (*e.g.*, top-7 or top-9) introduce irrelevant representations. Top-3 and top-5 provide the best balance, with top-5 achieving the strongest overall performance under multimodal settings.

Generalizability Analysis. To further validate the general applicability of our framework, we replace the backbone with an LLM-based pipeline, following [9, 26]. As illustrated in Fig. 2(c), RMoE consistently improves performance across all modality configurations, even when integrated with the LLM-based backbone. This demonstrates that our approach is not tied to a specific architecture, but can be effectively incorporated into alternative backbones.

4 Conclusion and Future Work

This paper presents Retrieval-based Mixture-of-Experts (RMoE), a novel framework for cancer survival prediction that effectively addresses the challenge of incomplete multimodal data. By integrating a bidirectional cross-modal retrieval with a reliability-aware MoE architecture, RMoE adaptively imputes missing information by capturing patient-specific cross-modal dependencies while main-

taining stable predictions across diverse missingness patterns. Experimental results across multiple datasets demonstrate that RMoE significantly outperforms state-of-the-art methods. In future work, we aim to extend our retrieval mechanism to larger external clinical cohorts and enhance the interpretability for deeper biological insights into patient-specific survival analysis.

Acknowledgments. This work was supported by grants from the Institute of Information & Communications Technology Planning & Evaluation (IITP), specifically the Artificial Intelligence Graduate School Program at Korea University (No. RS-2019-II190079), the National Research Foundation of Korea (NRF) funded by the Korea government (MSIT) (Nos. RS-2023-00212498, RS-2025-25302986, and RS-2025-25441950), and the Korean ARPA-H Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (No. RS-2025-25455839).

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

References

1. Adam, N., Wieder, R.: AI survival prediction modeling: The importance of considering treatments and changes in health status over time. Cancers **16**(20), 3527 (2024)
2. Cao, B., Bi, Z., Hu, Q., Zhang, H., Wang, N., Gao, X., Shen, D.: Autoencoder-driven multimodal collaborative learning for medical image synthesis. International Journal of Computer Vision **131**(8), 1995–2014 (2023)
3. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F., Jaume, G., Song, A.H., Chen, B., Zhang, A., Shao, D., Shaban, M., et al.: Towards a general-purpose foundation model for computational pathology. Nature Medicine **30**(3), 850–862 (2024)
4. 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)
5. 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)
6. Jaume, G., Vaidya, A., Chen, R.J., Williamson, D.F., Liang, P.P., Mahmood, F.: Modeling dense multimodal interactions between biological pathways and histology for survival prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11579–11590 (2024)
7. Klambauer, G., Unterthiner, T., Mayr, A., Hochreiter, S.: Self-normalizing neural networks. Advances in Neural Information Processing Systems **30** (2017)
8. Lee, C., Zame, W., Yoon, J., Van Der Schaar, M.: Deephit: A deep learning approach to survival analysis with competing risks. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
9. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C.H., Kang, J.: BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics **36**(4), 1234–1240 (2020)

10. Lee, Y.L., Tsai, Y.H., Chiu, W.C., Lee, C.Y.: Multimodal prompting with missing modalities for visual recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14943–14952 (2023)
11. Lim, M., Kang, B., Kam, T.E.: Multi-modal contrastive learning for tumor-specific missing modality synthesis. arXiv preprint arXiv:2502.19390 (2025)
12. 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)
13. Liu, F., Cai, L., Wang, Z., Fan, Z., Yu, J.g., Chen, H., Zhang, Y.: Prototype-guided cross-modal knowledge enhancement for adaptive survival prediction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 522–532. Springer (2025)
14. Liu, H., Wei, D., Lu, D., Sun, J., Wang, L., Zheng, Y.: M3AE: Multimodal representation learning for brain tumor segmentation with missing modalities. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 1657–1665 (2023)
15. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature Biomedical Engineering **5**(6), 555–570 (2021)
16. Meng, M., Gu, B., Fulham, M., Song, S., Feng, D., Bi, L., Kim, J.: Adaptive segmentation-to-survival learning for survival prediction from multi-modality medical images. NPJ Precision Oncology **8**(1), 232 (2024)
17. Ning, Z., Zhao, Z., Feng, Q., Chen, W., Xiao, Q., Zhang, Y.: Mutual-assistance learning for standalone mono-modality survival analysis of human cancers. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(6), 7577–7594 (2022)
18. Payne, K., Gavan, S.P., Wright, S.J., Thompson, A.J.: Cost-effectiveness analyses of genetic and genomic diagnostic tests. Nature Reviews Genetics **19**(4), 235–246 (2018)
19. Raser, J.M., O'shea, E.K.: Noise in gene expression: origins, consequences, and control. Science **309**(5743), 2010–2013 (2005)
20. Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., et al.: TransMIL: Transformer based correlated multiple instance learning for whole slide image classification. Advances in Neural Information Processing Systems **34**, 2136–2147 (2021)
21. Wang, H., Chen, Y., Ma, C., Avery, J., Hull, L., Carneiro, G.: Multi-modal learning with missing modality via shared-specific feature modelling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15878–15887 (2023)
22. Wang, H., Ma, C., Zhang, J., Zhang, Y., Avery, J., Hull, L., Carneiro, G.: Learnable cross-modal knowledge distillation for multi-modal learning with missing modality. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 216–226. Springer (2023)
23. Wang, Z., Zhang, Y., Xu, Y., Imoto, S., Chen, H., Song, J.: Histo-genomic knowledge association for cancer prognosis from histopathology whole slide images. IEEE Transactions on Medical Imaging **44**(5), 2170–2181 (2025)
24. Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M.: The cancer genome atlas pan-cancer analysis project. Nature Genetics **45**(10), 1113–1120 (2013)
25. Xu, Y., Chen, H.: Multimodal optimal transport-based co-attention transformer with global structure consistency for survival prediction. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 21241–21251 (2023)

26. Xu, Y., Zhou, F., Zhao, C., Wang, Y., Yang, C., Chen, H.: Distilled prompt learning for incomplete multimodal survival prediction. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 5102–5111 (2025)
27. Yang, H., Sun, J., Xu, Z.: Learning unified hyper-network for multi-modal mr image synthesis and tumor segmentation with missing modalities. IEEE Transactions on Medical Imaging **42**(12), 3678–3689 (2023)
28. Yu, C.N., Greiner, R., Lin, H.C., Baracos, V.: Learning patient-specific cancer survival distributions as a sequence of dependent regressors. Advances in neural information processing systems **24** (2011)
29. Zadeh, S.G., Schmid, M.: Bias in cross-entropy-based training of deep survival networks. IEEE Transactions on Pattern Analysis and Machine Intelligence **43**(9), 3126–3137 (2020)
30. Zhang, C., Chu, X., Ma, L., Zhu, Y., Wang, Y., Wang, J., Zhao, J.: M^3Care: Learning with missing modalities in multimodal healthcare data. In: Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining. pp. 2418–2428 (2022)
31. Zhang, Z., Chen, P., McGough, M., Xing, F., Wang, C., Bui, M., Xie, Y., Sapkota, M., Cui, L., Dhillon, J., et al.: Pathologist-level interpretable whole-slide cancer diagnosis with deep learning. Nature Machine Intelligence **1**(5), 236–245 (2019)