

Evidential Fusion Network for Multimodal Survival Prediction under Missing Modalities

Yucheng Xing^{1,2*}, Hailan Mo^{1*}, Zi Wang³, Ling Huang^{1,3†}, and Mengling Feng¹

¹ National University of Singapore, Singapore

² National University of Singapore Guangzhou Research Translation and Innovation Institute, Guangzhou, Guangdong, China

³ Imperial College London, London, UK

Abstract. Recent multimodal survival prediction models have demonstrated strong predictive performance by leveraging complementary information across modalities. However, such models generally assume data completeness and exhibit limited robustness toward missing modalities, which are frequently encountered in real-world clinical settings. We propose the Evidential Missing Modality Survival Fusion model for multimodal survival prediction under missing modalities. EMMS offers a straightforward, computationally effective approach to survival analysis without requiring a generative phase for missing data. By employing Dempster-Shafer theory and the newly introduced Gaussian Random Fuzzy Numbers for multimodal decision fusion, it considers both aleatoric and epistemic uncertainty alongside modality reliability for fusion. Moreover, the model treats missing modalities as vacuous evidence, preventing interference with available inputs and naturally reflecting increased uncertainty and calibrated predictions. Extensive experiments on four cancer datasets demonstrate state-of-the-art performance while providing calibrated and interpretable uncertainty estimates under incomplete multimodal observations, without extra computational overhead. Code is available at <https://github.com/YuchengXing99/EMMS>.

Keywords: Missing Modalities · Multimodal learning · Survival Analysis · Evidential Fusion · Gaussian Random Fuzzy Numbers · Uncertainty

1 Introduction

Accurate survival prediction is central to precision oncology, enabling personalized treatment planning and risk stratification. [20, 26, 28] Although many state-of-the-art multimodal survival analysis methods, e.g., MCAT [2], MOTCat [29], EsurvFusion [17], have provided more accurate predictions, the deployment of multimodal survival models in real-world clinical settings remains fundamentally constrained by missing modalities. In practice, a complete morpho-molecular

* Equal contribution.

† Corresponding author: Ling Huang (my@linghuang.co.uk).

profile that combines whole-slide images (WSIs) and transcriptomic measurements is often unavailable due to tissue exhaustion, sequencing costs, or heterogeneous digital pathology infrastructure across centers.

Imputation-based approaches, such as SMIL [21] and M3Care [32], reconstruct missing modalities through generative modeling or latent-space retrieval. Although effective in improving average performance, these methods risk hallucinating non-existent biological or morphological signals and often incur high computational cost, which is problematic in high-stakes clinical decision-making. In contrast, imputation-free approaches, such as EmbraceNet [4], aim for graceful degradation by aggregating only available embeddings. However, most treat missingness as a masking or dropout issue, without explicitly quantifying the information loss induced by absent modalities. The above research also exposes a critical uncertainty gap in multimodal survival analysis [16]. Clinical decision support requires distinguishing aleatoric uncertainty (intrinsic data noise) from epistemic uncertainty (lack of knowledge due to missing information), while most existing models provide only a “best guess,” without signaling their own ignorance.

To bridge this gap, we propose Evidential Missing Modality Survival Fusion (EMMS), a computationally efficient framework for multimodal survival prediction under missing modalities. EMMS is grounded in Dempster-Shafer (DS) evidential theory and represents each modality’s prediction as survival evidence parameterized by Gaussian Random Fuzzy Numbers (GRFNs) [6, 27]. Different from existing methods, **we bypass the complexities of data reconstruction and heuristic degradation**; instead, we treat missing modalities as vacuous evidence by setting their epistemic strength to zero. This formulation leads to a closed-form, evidence-weighted aggregation where available modalities contribute proportionally to their reliability, while missing inputs exert no influence. Consequently, the total evidential strength serves as a built-in, interpretable metric of the epistemic uncertainty induced by incomplete data. By unifying interaction-aware multimodal representation learning with principled evidential fusion, EMMS produces calibrated survival predictions that degrade gracefully under missing input conditions.

2 Related Work

Multimodal Survival Analysis. The integration of digital pathology and transcriptomics has enabled survival prediction models that capture complementary morphological and molecular information [8]. Recent methods emphasize interaction-based fusion. MCAT aligns genomic representations with discriminative WSI patches via co-attention [2], while MOTCat enforces global structural consistency using optimal transport [29]. SurvPath further incorporates biological priors by representing transcriptomics as pathway tokens and modeling dense pathway-patch interactions with a multimodal Transformer [19]. Despite strong performance, these architectures assume complete modality availability, limiting robustness in clinical cohorts where WSI or RNA data are frequently missing.

Learning with Missing Modalities. Handling missing modalities generally follows two primary paradigms: imputation-based or imputation-free learning. **Imputation-based** approaches reconstruct missing inputs through generative modeling or latent-space completion [1, 3, 10]. For example, SMIL employs prototype-weighted reconstruction with uncertainty regularization [21], and M3Care retrieves task-relevant neighbors to compensate for absent modalities [32]. While improving average performance, these methods may hallucinate spurious biological or morphological signals, raising concerns in high-stakes clinical settings. **Imputation-free** methods instead aggregate only observed inputs [11, 31], for example, EmbraceNet introduces stochastic embracement for partial embeddings [4]. However, most treat missingness as masking or dropout, without explicitly modeling the epistemic uncertainty caused by absent evidence, often resulting in miscalibrated confidence.

Evidential Survival Learning. Evidential learning framework based on Dempster-Shafer theory provides uncertainty-aware prediction by representing outputs as accumulated evidence [5, 7, 15, 23]. ENNreg first introduces Gaussian Random Fuzzy Numbers (GRFNs) for survival regression, separating aleatoric and epistemic uncertainty [6, 18]. EsurvFusion extends this framework to multi-modal fusion with learnable reliability discounting [17]. However, those models typically assume complete modalities. In this paper, we treat a missing modality as contributing vacuous evidence and naturally reducing total evidential strength in aggregations.

3 Method

3.1 Preliminaries

Gaussian Random Fuzzy Numbers. In epistemic random fuzzy sets theory, a GRFN is a random fuzzy subset of the real line defined by the membership function $\varphi(x; M, h) = \exp\left(-\frac{h}{2}(x - M)^2\right)$, whose mode M is a Gaussian random variable with $M \sim \mathcal{N}(\mu, \sigma^2)$. A GRFN can be represented by $\tilde{Y} \sim \tilde{\mathcal{N}}(\mu, \sigma^2, h)$, where μ is the location parameter, and σ and $h \in [0, +\infty)$ represent the aleatoric and epistemic uncertainties, respectively. Given a GRFN $\tilde{Y} \sim \tilde{\mathcal{N}}(\mu, \sigma^2, h)$, the degrees of plausibility and belief for any interval $[x, y]$ can be calculated using:

$$Pl_{\tilde{Y}}([x, y]) = \Phi\left(\frac{y-\mu}{\sigma}\right) - \Phi\left(\frac{x-\mu}{\sigma}\right) + pl_{\tilde{Y}}(x) \Phi\left(\frac{x-\mu}{\sigma_h}\right) + pl_{\tilde{Y}}(y) \left[1 - \Phi\left(\frac{y-\mu}{\sigma_h}\right)\right], \quad (1a)$$

$$Bel_{\tilde{Y}}([x, y]) = \Phi\left(\frac{y-\mu}{\sigma}\right) - \Phi\left(\frac{x-\mu}{\sigma}\right) - pl_{\tilde{Y}}(x) \left[\Phi\left(\frac{(x+y)/2 - \mu + (y-x)h\sigma^2/2}{\sigma_h}\right) - \Phi\left(\frac{x-\mu}{\sigma_h}\right) \right] + pl_{\tilde{Y}}(y) \left[\Phi\left(\frac{(x+y)/2 - \mu - (y-x)h\sigma^2/2}{\sigma_h}\right) - \Phi\left(\frac{y-\mu}{\sigma_h}\right) \right], \quad (1b)$$

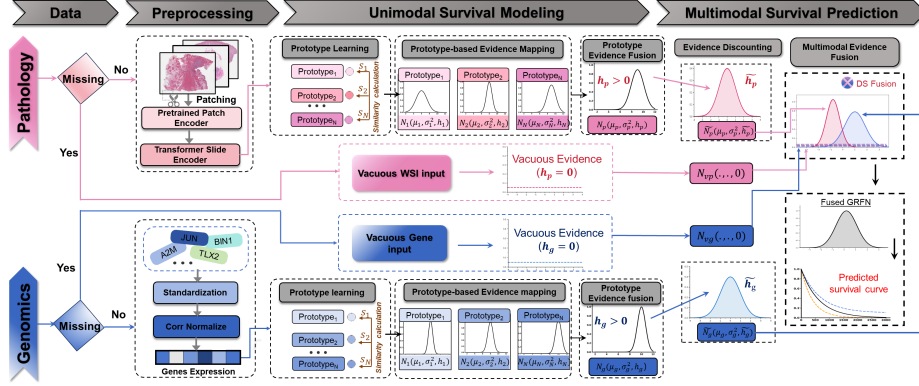


Fig. 1. Overview of the proposed EMMS framework. WSI and gene modalities generate GRFN survival evidence, modeling both aleatoric and epistemic uncertainty. Missing modalities are treated as vacuous evidence and fused via Dempster–Shafer combination to produce calibrated survival predictions.

where $\sigma_h = \sigma\sqrt{1+h\sigma^2}$ and $pl_{\tilde{Y}}(x) = \frac{1}{\sqrt{1+h\sigma^2}} \exp\left(-\frac{h(x-\mu)^2}{2(1+h\sigma^2)}\right)$ is the contour function and Φ denotes the standard normal cumulative distribution function.

Vacuous Evidence. In the presence of missing modalities, the corresponding source provides no informative evidence. In the GRFN framework, this situation is naturally modeled by setting $h = 0$. In this case, the epistemic component vanishes and the GRFN degenerates into vacuous evidence, thereby representing complete ignorance.

GRFNs for Survival Modeling. Let $T > 0$ denote survival time and define $Y = \log T$ to map the positive domain onto the real line. Under a GRFN model $\tilde{Y}$, the survival function $S(t) = \mathbb{P}(T > t)$ is bounded as

$$Bel_{\tilde{Y}}((\log t, \infty)) \leq S(t) \leq Pl_{\tilde{Y}}((\log t, \infty)). \quad (2)$$

3.2 Unimodal Survival Modeling

Pathology prediction branch. Given a WSI, we tessellate it into N_p patches $\{\mathbf{x}_{i,p}\}_{i=1}^{N_p}$ at 20x magnification. Each patch is encoded by a pretrained patch encoder f_{Patch} , yielding patch embeddings $\mathbf{h}_{i,p} = f_{\text{Patch}}(\mathbf{x}_{i,p})$. The resulting set of patch features is then processed by the TITAN slide encoder [9], which captures contextual dependencies among patches via a transformer-based architecture and aggregates them into a slide-level representation:

$$\mathbf{z}_p = f_{\text{TITAN}}(\{\mathbf{h}_{i,p}\}_{i=1}^{N_p}), \quad (3)$$

where $\mathbf{z}_p \in \mathbb{R}^D$ serves as the global WSI feature for time-to-event modeling.

Following a prototype-based evidential time-to-event formulation [18], we map an input feature vector $\mathbf{z}_p$ to a survival evidence representation. Given K prototypes $\{\mathbf{w}_k\}_{k=1}^K$, the prototype embedding layer evaluates the affinity between $\mathbf{z}_p$ and each prototype via a radial basis function kernel with a positive scale parameter γ_k :

$$s_k(\mathbf{z}_p) = \exp(-\gamma_k^2(\mathbf{z}_p - \mathbf{w}_k)^\top \text{Diag}(\mathbf{v}_p)(\mathbf{z}_p - \mathbf{w}_k)), \quad (4)$$

where $\mathbf{v}_p \in \mathbb{R}_+^D$ is a learnable modality-specific scaling vector defining a diagonal Mahalanobis metric for adaptive dimension-wise reweighting. For each prototype $\mathbf{w}_k$, the evidence mapping layer produces a GRFN

$$\tilde{Y}_k(\mathbf{z}_p) \sim \tilde{\mathcal{N}}(\mu_k(\mathbf{z}_p), \sigma_k^2, s_k(\mathbf{z}_p) h_k), \quad (5)$$

where σ_k^2 and h_k denote the prototype-specific variance and precision, respectively. The location parameter is modeled linearly as

$$\mu_k(\mathbf{z}_p) = \beta_k^\top \mathbf{z}_p + \beta_{k0}, \quad (6)$$

with β_k and β_{k0} being learnable parameters. Finally, evidence from all prototypes is integrated by the generalized product–intersection combination rule, yielding the aggregated GRFN $\tilde{Y}_p(\mathbf{z}_p) \sim \tilde{\mathcal{N}}(\mu_p(\mathbf{z}_p), \sigma_p^2(\mathbf{z}_p), h_p(\mathbf{z}_p))$, where

$$\mu_p(\mathbf{z}_p) = \frac{\sum_{k=1}^K s_k(\mathbf{z}_p) h_k \mu_k(\mathbf{z}_p)}{\sum_{k=1}^K s_k(\mathbf{z}_p) h_k}, \quad \sigma_p^2(\mathbf{z}_p) = \frac{\sum_{k=1}^K s_k^2(\mathbf{z}_p) h_k^2 \sigma_k^2}{\left(\sum_{k=1}^K s_k(\mathbf{z}_p) h_k\right)^2}, \quad (7)$$

and $h_p(\mathbf{z}_p) = \sum_{k=1}^K s_k(\mathbf{z}_p) h_k$. Here $\mu_p(\mathbf{z}_p)$ corresponds to the most plausible event-time estimate, $\sigma_p^2(\mathbf{z}_p)$ captures aleatoric uncertainty from data variability, and $h_p(\mathbf{z}_p)$ characterizes epistemic uncertainty due to model ignorance.

Genomic prediction branch. For each individual, transcriptomic profile includes N_g gene expression measurements and is represented as $\mathbf{z}_g = \{x_{i,g}\}_{i=1}^{N_g}$, where $x_{i,g} \in \mathbb{R}$. Following the same prediction framework as the pathology branch, the genomic branch produces survival evidence modeled as a GRFN

$$\tilde{Y}_g(\mathbf{z}_g) \sim \tilde{\mathcal{N}}(\mu_g(\mathbf{z}_g), \sigma_g^2(\mathbf{z}_g), h_g(\mathbf{z}_g)).$$

3.3 Evidential Fusion under Missing Modalities

Let $\tilde{Y}_p(\mathbf{z}_p)$ and $\tilde{Y}_g(\mathbf{z}_g)$ denote the modality-specific survival evidence derived from pathology and genomics, respectively. We fuse these evidences using Dempster’s combination rule within the Dempster–Shafer framework, enabling principled aggregation of uncertainty.

To explicitly handle missing modalities, we introduce a binary availability indicator $r \in \{0, 1\}$ for each modality. The corresponding GRFN is defined as

$$\tilde{Y}^{(r)}(\mathbf{z}) = \begin{cases} \tilde{\mathcal{N}}(\mu(\mathbf{z}), \sigma^2(\mathbf{z}), h(\mathbf{z})), & r = 1, \\ \tilde{\mathcal{N}}(\cdot, \cdot, 0), & r = 0, \end{cases} \quad (8)$$

where $h(\mathbf{z})$ denotes the epistemic strength of evidence. Setting $h = 0$ corresponds to vacuous evidence, representing complete ignorance in the DS sense. Importantly, this formulation does not approximate or reconstruct missing inputs; instead, it encodes their absence as zero evidential contribution.

Given two modalities with epistemic strengths h_p and h_g , the fused GRFN $\tilde{Y}_f \sim \tilde{\mathcal{N}}(\mu_f, \sigma_f^2, h_f)$ is obtained as

$$\mu_f = \frac{d(h_p)\mu_p + d(h_g)\mu_g}{d(h_p) + d(h_g)}, \quad \sigma_f^2 = \frac{d(h_p)^2\sigma_p^2 + d(h_g)^2\sigma_g^2}{(d(h_p) + d(h_g))^2}, \quad h_f = d(h_p) + d(h_g), \quad (9)$$

where $d(\cdot)$ denotes a monotonic discounting function defined as $d(h) = \rho \log(1 + h)$, with $\rho \in (0, 1)$ controlling the discounting intensity. The logarithmic form ensures bounded growth and diminishing sensitivity for large epistemic values, which stabilizes fusion under heterogeneous uncertainty levels.

The evidential fusion operation yields several desirable properties. First, fusion is evidence-weighted: modalities with higher epistemic strength contribute more to the fused prediction. Second, the total epistemic strength accumulates additively, providing an interpretable quantitative measure of overall confidence. Crucially, when a modality is missing ($r = 0$), its epistemic parameter vanishes ($h = 0$), and the fusion naturally reduces to the available modality: $(\mu_f, \sigma_f^2, h_f) = (\mu_{available}, \sigma_{available}^2, d(h_{available}))$. Thus, missing modalities contribute vacuous evidence and do not interfere with observed data. Meanwhile, a reduction in total evidential strength explicitly reflects missingness-induced epistemic uncertainty.

3.4 Loss function

We follow the generalized negative log-likelihood for GRFN-based time-to-event learning [16]. The survival loss $\mathcal{L}_{\text{surv}}$ combines belief- and plausibility-based terms to handle both uncensored and right-censored samples: uncensored cases supervise a small interval around $y = \log(t)$, while censored cases supervise the tail event $[y, \infty)$.

Let $\tilde{Y}_p, \tilde{Y}_g$ be the unimodal evidence and $\tilde{Y}_f$ the fused evidence. With modality availability flags $r_p, r_g \in \{0, 1\}$, our objective is

$$\mathcal{L} = r_p \mathcal{L}_{\text{surv}}^p + r_g \mathcal{L}_{\text{surv}}^g + \lambda_{\text{fus}} \mathbb{I}[r_p \vee r_g] \mathcal{L}_{\text{surv}}^f + \lambda_{\text{align}} \mathbb{I}[r_p \wedge r_g] \mathcal{L}_{\text{align}}, \quad (10)$$

where the complete-case alignment term is defined as

$$\mathcal{L}_{\text{align}} = \text{KL}(\mathcal{N}(\mu_p, \sigma_p^2) \parallel \mathcal{N}(\mu_g, \sigma_g^2)) + \text{KL}(\mathcal{N}(\mu_g, \sigma_g^2) \parallel \mathcal{N}(\mu_p, \sigma_p^2)). \quad (11)$$

where $\text{KL}(\cdot \parallel \cdot)$ denotes Kullback–Leibler divergence, and the symmetric KL between two Gaussians is used as a lightweight cross-modal consistency regularizer.

Table 1. Results under **0% missing modality**. C : C-index $\uparrow$, S : IBS $\downarrow$. **G**, **P**, and **C** denote genomics-only, pathology-only, and complete-modality evaluation. Best results are in **bold**, and our method is highlighted in **blue**.

Method	BRCA		LUAD		STAD		KIRC		
	C ↑	S ↓	C ↑	S ↓	C ↑	S ↓	C ↑	S ↓	
G	DisPro	0.511±0.17	0.100±0.03	0.587±0.04	0.282±0.09	0.542±0.11	0.261±0.03	0.664±0.03	0.145±0.04
	MUSE	0.661±0.10	0.134±0.07	0.586±0.03	0.282±0.03	0.472±0.09	0.299±0.04	0.664±0.11	0.168±0.07
	Flex-MoE	0.582±0.08	0.068±0.04	0.579±0.05	0.150±0.04	0.598±0.06	0.172±0.06	0.552±0.12	0.104±0.03
	Ours	0.725±0.09	0.060±0.04	0.589±0.07	0.145±0.04	0.517±0.08	0.171±0.07	0.685±0.12	0.101±0.04
P	DisPro	0.641±0.19	0.104±0.06	0.600±0.11	0.179±0.05	0.539±0.09	0.206±0.06	0.639±0.15	0.150±0.05
	MUSE	0.665±0.06	0.104±0.09	0.611±0.04	0.153±0.05	0.563±0.11	0.175±0.07	0.818±0.07	0.099±0.03
	Flex-MoE	0.717±0.09	0.153±0.08	0.635±0.06	0.212±0.07	0.515±0.09	0.232±0.09	0.783±0.07	0.151±0.05
	Ours	0.720±0.10	0.057±0.03	0.638±0.08	0.139±0.04	0.586±0.10	0.170±0.07	0.801±0.06	0.093±0.03
C	DisPro	0.690±0.09	0.076±0.04	0.583±0.05	0.207±0.05	0.543±0.11	0.252±0.03	0.631±0.16	0.140±0.05
	MUSE	0.682±0.09	0.099±0.07	0.601±0.03	0.165±0.04	0.488±0.09	0.209±0.05	0.732±0.10	0.111±0.05
	MOTCat	0.619±0.11	0.118±0.06	0.611±0.02	0.199±0.04	0.554±0.06	0.224±0.06	0.742±0.14	0.106±0.03
	Flex-MoE	0.723±0.07	0.066±0.04	0.612±0.07	0.160±0.03	0.499±0.09	0.201±0.07	0.768±0.05	0.100±0.03
	Ours	0.740±0.08	0.057±0.03	0.655±0.07	0.143±0.04	0.555±0.09	0.174±0.07	0.801±0.08	0.098±0.03

4 Experiments

4.1 Experimental Setup

Datasets. Four cancer cohorts from The Cancer Genome Atlas (TCGA) are evaluated in this study, including Breast Invasive Carcinoma (BRCA), Stomach Adenocarcinoma (STAD), Kidney Renal Clear Cell Carcinoma (KIRC), and Lung Adenocarcinoma (LUAD). To prevent the model from exploiting center-specific artifacts such as staining patterns or scanner variations, we employ a 5-fold site-stratified cross-validation strategy [14].

Baselines. We employ DisPro [30] and KNN [22] as an imputation-based baseline, and MUSE [25] and Flex-Moe [31] as an imputation-free baseline, and additionally consider MOTCat [29] in the no-missingness scenario. Following prior works [21, 24], we simulate missing modalities by randomly masking fully paired samples under a total 60% missing rate, yielding five pathology/genomics splits: (0%,60%), (20%,40%), (30%,30%), (40%,20%), and (60%,0%). For each missing setting, evaluation is conducted under pathology-only, genomics-only, and complete-modality settings. For EMMS, we set $K = 50$ for each modality branch and initialize the prototypes deterministically using k-means clustering on the training features to promote diverse coverage of the feature space. Prototype precisions h_k are initialized identically across branches, placing their epistemic strengths on a common scale. The fused survival loss jointly updates h_p and h_g , preserving their cross-branch comparability during training.

Evaluation Metrics. Model discrimination performance is assessed using the Concordance index (C-index) [13]. Calibration performance is evaluated using the integrated Brier score (IBS) [12].

4.2 Performance comparison under different missing scenarios

Table 1 reports the results under the 0% missing-modality setting. Across the four datasets, our method achieves the best performance in most cases. Com-

Table 2. Results under **60% missing modality**. C : C-index $\uparrow$, S : IBS $\downarrow$. **G**, **P**, and **C** denote genomics-only, pathology-only, and complete-modality test cases. Best results are in **bold**, and our method is highlighted in **blue**.

Method	BRCA		LUAD		STAD		KIRC	
	$C \uparrow$	$S \downarrow$	$C \uparrow$	$S \downarrow$	$C \uparrow$	$S \downarrow$	$C \uparrow$	$S \downarrow$
KNN	0.574 $\pm$ 0.06	0.091 $\pm$ 0.04	0.540 $\pm$ 0.05	0.190 $\pm$ 0.04	0.521 $\pm$ 0.04	0.210 $\pm$ 0.07	0.574 $\pm$ 0.06	0.105 $\pm$ 0.04
DisPro	0.639 $\pm$ 0.12	0.089 $\pm$ 0.03	0.550 $\pm$ 0.07	0.250 $\pm$ 0.06	0.522 $\pm$ 0.07	0.259 $\pm$ 0.06	0.616 $\pm$ 0.10	0.190 $\pm$ 0.08
MUSE	0.650 $\pm$ 0.08	0.114 $\pm$ 0.09	0.589 $\pm$ 0.06	0.176 $\pm$ 0.03	0.514 $\pm$ 0.12	0.229 $\pm$ 0.07	0.607 $\pm$ 0.18	0.137 $\pm$ 0.07
Flex-MoE	0.608 $\pm$ 0.07	0.108 $\pm$ 0.09	0.573 $\pm$ 0.02	0.165 $\pm$ 0.01	0.524 $\pm$ 0.04	0.189 $\pm$ 0.02	0.624 $\pm$ 0.06	0.104 $\pm$ 0.02
Ours	0.693$\pm$0.11	0.060$\pm$0.04	0.603$\pm$0.09	0.144$\pm$0.04	0.530$\pm$0.10	0.174$\pm$0.07	0.648$\pm$0.13	0.102$\pm$0.04
KNN	0.570 $\pm$ 0.08	0.120 $\pm$ 0.07	0.525 $\pm$ 0.08	0.211 $\pm$ 0.08	0.521 $\pm$ 0.05	0.259 $\pm$ 0.08	0.617 $\pm$ 0.08	0.120 $\pm$ 0.05
DisPro	0.567 $\pm$ 0.13	0.138 $\pm$ 0.11	0.569 $\pm$ 0.06	0.189 $\pm$ 0.03	0.528 $\pm$ 0.10	0.239 $\pm$ 0.08	0.755 $\pm$ 0.09	0.126 $\pm$ 0.08
MUSE	0.647 $\pm$ 0.10	0.142 $\pm$ 0.13	0.609 $\pm$ 0.07	0.180 $\pm$ 0.03	0.567 $\pm$ 0.08	0.216 $\pm$ 0.06	0.779 $\pm$ 0.07	0.104 $\pm$ 0.03
Flex-MoE	0.594 $\pm$ 0.05	0.083 $\pm$ 0.04	0.595 $\pm$ 0.01	0.167 $\pm$ 0.04	0.476 $\pm$ 0.02	0.226 $\pm$ 0.05	0.753 $\pm$ 0.02	0.103 $\pm$ 0.01
Ours	0.686$\pm$0.09	0.059$\pm$0.03	0.631$\pm$0.07	0.139$\pm$0.04	0.611$\pm$0.08	0.168$\pm$0.07	0.793$\pm$0.08	0.097$\pm$0.03
KNN	0.584 $\pm$ 0.07	0.068 $\pm$ 0.04	0.537 $\pm$ 0.04	0.164 $\pm$ 0.04	0.535 $\pm$ 0.04	0.212 $\pm$ 0.07	0.611 $\pm$ 0.09	0.103 $\pm$ 0.03
DisPro	0.575 $\pm$ 0.10	0.095 $\pm$ 0.05	0.576 $\pm$ 0.05	0.208 $\pm$ 0.05	0.530 $\pm$ 0.06	0.229 $\pm$ 0.06	0.648 $\pm$ 0.12	0.135 $\pm$ 0.06
MUSE	0.698 $\pm$ 0.08	0.105 $\pm$ 0.09	0.624 $\pm$ 0.05	0.154 $\pm$ 0.03	0.541 $\pm$ 0.09	0.191 $\pm$ 0.05	0.737 $\pm$ 0.13	0.104 $\pm$ 0.04
Flex-MoE	0.605 $\pm$ 0.08	0.070 $\pm$ 0.01	0.606 $\pm$ 0.01	0.161 $\pm$ 0.01	0.481 $\pm$ 0.04	0.196 $\pm$ 0.01	0.713 $\pm$ 0.03	0.103 $\pm$ 0.02
Ours	0.715$\pm$0.06	0.059$\pm$0.04	0.637$\pm$0.07	0.142$\pm$0.04	0.576$\pm$0.11	0.174$\pm$0.07	0.753$\pm$0.10	0.102$\pm$0.04

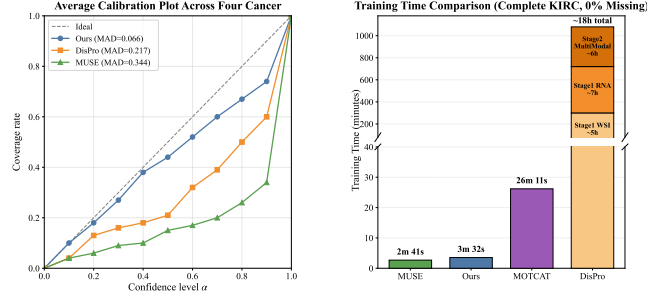


Fig. 2. Left: calibration curves (MAD, mean absolute deviation from the diagonal used to quantify calibration). Right: training time comparison across methods.

pared with the baselines, it generally attains higher C-index and lower IBS, indicating improved discrimination and calibration under fully observed conditions. Table 2 presents the results under the 60% missing-modality setting. Across the four datasets, our method achieves the best performance in all cases and demonstrates strong robustness under substantial modality absence. Overall, these results demonstrate the effectiveness and robustness of our framework across both complete and incomplete modality scenarios.

4.3 Calibration Analysis and Computational Cost

Figure 2 shows that our method remains closer to the diagonal line across all four datasets, indicating better calibration than the compared baselines. This consistent behavior suggests improved confidence estimation and highlights the reliability of our uncertainty quantification in survival prediction. In terms of computational cost, our approach is comparable to the imputation-free baseline

MUSE, while remaining substantially more efficient than the imputation-based method DisPro, indicating that EMMS achieves strong robustness and calibration performance without introducing additional computational overhead.

5 Conclusion

In this paper, we introduced EMMS, a survival analysis framework for incomplete multimodal data. Leveraging Dempster–Shafer theory and GRFNs, it models missing modalities as vacuous evidence without introducing extra computational cost, preventing fusion contamination while capturing increased predictive uncertainty under missingness. Experiments on four public cancer datasets show SOTA discrimination and calibration across diverse clinical cohorts and varying missing-modality patterns, standing as a scalable, computation-effective, uncertainty-aware, and theoretically grounded solution for personalized medicine in realistic clinical settings.

Acknowledgments. This research was supported by the National Medical Research Council (NMRC) Healthy and Meaningful Longevity–Cognition Grant (NICCOG2024-0028), the AI for Public Health Program at the Saw Swee Hock School of Public Health, National University of Singapore, and the Dame Julia Higgins Postdoc Collaborative Research Fund. This work was also supported by the NUS Guangzhou Research Translation and Innovation Institute (GRTII)-Affiliated PhD Scholarship Program.

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

References

1. Chen, J., Zhang, A.: Hgmf: heterogeneous graph-based fusion for multimodal data with incompleteness. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 1295–1305 (2020)
2. 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)
3. Chen, Z., Du, Y., Hu, J., Liu, Y., Li, G., Wan, X., Chang, T.H.: Multi-modal masked autoencoders for medical vision-and-language pre-training. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer (2022)
4. Choi, J.H., Lee, J.S.: Embracenet: A robust deep learning architecture for multimodal classification. *Information Fusion* **51**, 259–270 (2019)
5. Denceux, T.: Parametric families of continuous belief functions based on generalized gaussian random fuzzy numbers. *Fuzzy Sets and Systems* **471**, 108679 (2023)
6. Denceux, T.: Quantifying prediction uncertainty in regression using random fuzzy sets: the ennreg model. *IEEE Transactions on Fuzzy Systems* **31**(10), 3690–3699 (2023)

7. Denœux, T.: Reasoning with fuzzy and uncertain evidence using epistemic random fuzzy sets: General framework and practical models. *Fuzzy Sets and Systems* **453**, 1–36 (2023)
8. Ding, K., Zhou, M., Metaxas, D.N., Zhang, S.: Pathology-and-genomics multimodal transformer for survival outcome prediction. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 622–631. Springer (2023)
9. Ding, T., Wagner, S.J., Song, A.H., Chen, R.J., Lu, M.Y., Zhang, A., Vaidya, A.J., Jaume, G., Shaban, M., Kim, A., et al.: A multimodal whole-slide foundation model for pathology. *Nature Medicine* pp. 1–13 (2025)
10. Geng, X., Liu, H., Lee, L., Schuurmans, D., Levine, S., Abbeel, P.: Multimodal masked autoencoders learn transferable representations. *arXiv preprint arXiv:2205.14204* (2022)
11. Gong, L., Liu, Y., Sun, L., Bi, Y., Liu, J., Zhu, X.: Embracing aleatoric uncertainty in medical multimodal learning with missing modalities. *arXiv preprint arXiv:2601.21950* (2026)
12. Graf, E., Schmoor, C., Sauerbrei, W., Schumacher, M.: Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine* **18**(17–18), 2529–2545 (1999)
13. 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)
14. Howard, F.M., Dolezal, J., Kochanny, S., Schulte, J., Chen, H., Heij, L., Huo, D., Nanda, R., Olopade, O.I., Kather, J.N., et al.: The impact of site-specific digital histology signatures on deep learning model accuracy and bias. *Nature communications* **12**(1), 4423 (2021)
15. Huang, L., Ruan, S., Xing, Y., Feng, M.: A review of uncertainty quantification in medical image analysis: Probabilistic and non-probabilistic methods. *Medical Image Analysis* **97**, 103223 (2024)
16. Huang, L., Xing, Y., Denœux, T., Feng, M.: An evidential time-to-event prediction model based on gaussian random fuzzy numbers. In: *International Conference on Belief Functions*. pp. 49–57. Springer (2024)
17. Huang, L., Xing, Y., Lin, Q., Duan, J., Ruan, S., Feng, M.: Esurvfusion: An evidential multimodal survival fusion model based on epistemic random fuzzy sets. *IEEE Transactions on Fuzzy Systems* (2025)
18. Huang, L., Xing, Y., Mishra, S., Denœux, T., Feng, M.: Evidential time-to-event prediction with calibrated uncertainty quantification. *International Journal of Approximate Reasoning* **181**, 109403 (2025)
19. 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)
20. Liu, P., Zeng, X., Ma, T., Xing, Y., Ren, X., Liu, Y.: Sparse task vector mixup with hypernetworks for efficient knowledge transfer in whole-slide image prognosis. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 35238–35247 (2026)
21. Ma, M., Ren, J., Zhao, L., Tulyakov, S., Wu, C., Peng, X.: Smil: Multimodal learning with severely missing modality. In: *Proceedings of the AAAI conference on artificial intelligence*. vol. 35, pp. 2302–2310 (2021)
22. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: Machine learning in python. *the Journal of machine Learning research* **12**, 2825–2830 (2011)

23. Shafer, G.: Dempster-shafer theory. *Encyclopedia of artificial intelligence* 1(330-331), 3 (1992)
24. 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)
25. Wu, Z., Dadu, A., Tustison, N., Avants, B., Nalls, M., Sun, J., Faghri, F.: Multimodal patient representation learning with missing modalities and labels. In: *The Twelfth International Conference on Learning Representations* (2024), <https://openreview.net/forum?id=Je5SHCKpPa>
26. Xing, Y., Huang, L., Liu, P., Ma, J., Xu, J., He, K., Feng, M.: Semantic-anchored evidential fusion for domain-robust whole-slide survival analysis. *arXiv preprint arXiv:2606.19966* (2026)
27. Xing, Y., Huang, L., Ma, J., Hong, R., Qiu, J., Liu, P., He, K., Fu, H., Feng, M.: Dpsurv: Dual-prototype evidential fusion for uncertainty-aware and interpretable whole-slide image survival prediction. *arXiv preprint arXiv:2510.00053* (2025)
28. Xing, Y., Liu, P., Ma, J., Hong, R., Qiu, J., Liu, T., He, K., Huang, L., Feng, M.: Bridging the modality bottleneck in pathology mil through virtual molecular staining. *arXiv preprint arXiv:2605.16392* (2026)
29. 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)
30. Xu, Y., Zhou, F., Zhao, C., Wang, Y., Yang, C., Chen, H.: Distilled prompt learning for incomplete multimodal survival prediction. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 5102–5111 (June 2025)
31. Yun, S., Choi, I., Peng, J., Wu, Y., Bao, J., Zhang, Q., Xin, J., Long, Q., Chen, T.: Flex-moe: Modeling arbitrary modality combination via the flexible mixture-of-experts (2024), <https://arxiv.org/abs/2410.08245>
32. Zhang, C., Chu, X., Ma, L., Zhu, Y., Wang, Y., Wang, J., Zhao, J.: M3care: 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)